GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 21:22:46 ; Search time 8287 Seconds
(without alignments)
11386.242 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		2177	100.0	2177	6	BD269644	BD269644 Human sec
2		1994	91.6	2089	6	AX357116	AX357116 Sequence
3		1891	86.9	2681	9	HSA276171	AJ276171 Homo sapi
4		1759	80.8	2589	6	AX357100	AX357100 Sequence
5		1758	80.8	2039	6	AX357115	AX357115 Sequence
6		1758	80.8	2145	6	AX357117	AX357117 Sequence
7		1745	80.2	1959	6	AX470036	AX470036 Sequence
8		1745	80.2	2507	6	AX470035	AX470035 Sequence
9		1745	80.2	2507	9	HSA279016	AJ279016 Homo sapi
10		1744	80.1	1947	6	AX470056	AX470056 Sequence
11		1744	80.1	1947	9	HSA421515	AJ421515 Homo sapi
12		1687	77.5	2889	9	AK057190	AK057190 Homo sapi
13		1617	74.3	2413	6	AX427468	AX427468 Sequence
14		1438	66.1	2147	9	BC034245	BC034245 Homo sapi
15		1428	65.6	2263	6	AX202100	AX202100 Sequence
16		1314	60.4	1501	6	AR339541	AR339541 Sequence
17		1236	56.8	2178	6	AX876177	AX876177 Sequence
18		1236	56.8	2178	6	BD155992	BD155992 Primer fo
19		1236	56.8	2178	9	AK001182	AK001182 Homo sapi
20		738	33.9	789	6	AR379939	AR379939 Sequence
21		648	29.8	1143	6	BD205640	BD205640 97 human
22		496	22.8	608	6	AX867169	AX867169 Sequence
23		496	22.8	608	6	BD147231	BD147231 Primer fo
24	c	225	10.3	146124	9	AL358938	AL358938 Human DNA
25	c	203	9.3	161235	9	AL139239	AL139239 Human DNA
26	c	203	9.3	172383	2	AC015630	AC015630 Homo sapi
27		160	7.3	172383	2	AC015630	AC015630 Homo sapi
28		143	6.6	409	11	HSJ123C5	AL159715 STS from
29	c	113	5.2	395	11	G25373	G25373 human STS E
30	c	113	5.2	492	11	G27517	G27517 human STS S
31		109	5.0	707	6	AR379664	AR379664 Sequence
32		57	2.6	57	6	AX470037	AX470037 Sequence
33		54	2.5	1558	10	BC024472	BC024472 Mus muscu
34		54	2.5	1973	10	MMU421516	AJ421516 Mus muscu
35	c	54	2.5	163348	10	AC119236	AC119236 Mus muscu
36		54	2.5	189196	10	AL603804	AL603804 Mouse DNA
37		44	2.0	700	10	RNU78304	U78304 Rattus norv
38		40	1.8	752	6	AX202149	AX202149 Sequence
39	c	30	1.4	210385	10	AC106128	AC106128 Rattus no
40		30	1.4	243770	2	AC120763	AC120763 Rattus no
41	c	30	1.4	252404	2	AC097117	AC097117 Rattus no
42	c	29	1.3	164766	10	AC144792	AC144792 Mus muscu
43	c	25	1.1	25	6	AX357106	AX357106 Sequence
44		24	1.1	2012	8	AY237400	AY237400 Malus x d
45		23	1.1	23	6	AX357108	AX357108 Sequence

ALIGNMENTS

RESULT 1
BD269644
LOCUS BD269644 2177 bp DNA linear PAT 17-JUL-2003
DEFINITION Human secretory proteins.
ACCESSION BD269644
VERSION BD269644.1 GI:33079412
KEYWORDS JP 2002537805-A/13.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2177)
AUTHORS Tang,T.Y., Lal,P., Baughn,M.R., Yue,H., Young,J.A., Lu,D.A.M. and Azimzai,Y.
TITLE Human secretory proteins

```
  JOURNAL   Patent: JP 2002537805-A 13 12-NOV-2002;
            INCYTE PHARMACEUTICALS INC
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002537805-A/13
            PD   12-NOV-2002
            PF   03-MAR-2000 JP 2000602763
            PR   05-MAR-1999 US    60/123117
            PI   TOM Y TANG,PREETI LAL,MARIAH R BAUGHN,HENRY YUE,JANICE AU  PI
             YOUNG,
            PI   DYUNG AINA M LU,YALDA AZIMZAI
            PC   C12N15/09,A01K67/027,A61K38/00,A61K38/22,A61K45/00,A61P1/00,
            PC   A61P9/00,
            PC   A61P25/00,A61P29/00,A61P31/00,A61P35/00,C07K14/47,C07K16/18,
            PC   C12N1/15,
            PC   C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00 PC
            ,C12N5/00,A61K37/02,
            PC   A61K37/24
            CC   Incyte ID No: 2991370CB1
            FH   Key             Location/Qualifiers
            FT   source          1. .2177
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .2177
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 2177;  DB 6;  Length 2177;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2177;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560

Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Qy       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980

Qy       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040

Qy       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100

Qy       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160

Qy       2161 TCTTGAAAAAAAAAAAA 2177
             |||||||||||||||||
Db       2161 TCTTGAAAAAAAAAAAA 2177


RESULT 2
AX357116
LOCUS       AX357116                2089 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 17 from Patent WO0206478.
ACCESSION   AX357116
VERSION     AX357116.1  GI:18674296
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 17 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2089
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           91.6%;  Score 1994;  DB 6;  Length 2089;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2044;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304
```

```
Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1992
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1845 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1904

Qy       1993 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 2052
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1905 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 1964

Qy       2053 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2112
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1965 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2024

Qy       2113 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2025 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2084

Qy       2173 AAAAA 2177
             |||||
Db       2085 AAAAA 2089
```

```
RESULT 3
HSA276171
LOCUS       HSA276171              2681 bp    mRNA    linear   PRI 20-JUL-2000
DEFINITION  Homo sapiens mRNA for ASPIC (acidic secreted protein in cartilage)
            (ASPIC1 gene).
ACCESSION   AJ276171
VERSION     AJ276171.1  GI:9368805
KEYWORDS    acidic secreted protein in cartilage; ASPIC1 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bolton,M.C., Wait,R. and Saklatvala,J.
  TITLE     Cloning of ASPIC, a novel protein secreted by human normal and
            osteoarthritic cartilage, identified by 2D electrophoresis and mass
            spectrometry
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2681)
  AUTHORS   Bolton,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2000) Bolton M.C., Cell signalling, Kennedy
            Institute of Rheumatology, 1 Aspenlea Road, Hammersmith, London W6
            8LH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2681
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     gene            1. .2681
                     /gene="ASPIC1"
     5'UTR           1. .149
                     /gene="ASPIC1"
                     /note="A second transcription start site at +31"
                     /evidence=experimental
     CDS             150. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98267.1"
                     /db_xref="GI:9368806"
                     /db_xref="GOA:Q9NQ80"
                     /db_xref="SPTREMBL:Q9NQ80"
                     /translation="MAPSADPGMSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLP
                     PDYDSNPTQLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSS
                     PYYALRDRQGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDI
                     LSDEVNVARGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDL
                     SRGILALRDVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTF
                     VDAAASAGVDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDI
                     ASPKFSMPSPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEEL
                     NPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVP
                     RTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWP
                     DGKMVSRNVASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQF
                     PFVCPRDKPVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAA
                     ATAAAAAAAGAATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     150. .230
                     /gene="ASPIC1"
     CDS             174. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98268.1"
                     /db_xref="GI:9368807"
                     /db_xref="GOA:Q9NQ79"
                     /db_xref="SPTREMBL:Q9NQ79"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAA
                     AGAATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     174. .230
                     /gene="ASPIC1"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
```

```
     CDS             183. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98269.1"
                     /db_xref="GI:9368808"
                     /db_xref="GOA:Q9NQ78"
                     /db_xref="SPTREMBL:Q9NQ78"
                     /translation="MLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPTQLN
                     YGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDRQGN
                     AIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVARGV
                     ASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALRDVA
                     AEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAGVDD
                     PHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMPSPV
                     RTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEPEGR
                     GTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFARGA
                     KVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRNVAS
                     GEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDKPVC
                     VNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAAAGA
                     ATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     183. .230
                     /gene="ASPIC1"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
     mat_peptide     231. .2132
                     /gene="ASPIC1"
                     /product="ASPIC"
                     /standard_name="acidic secreted protein in cartilage"
     3'UTR           2136. .2681
                     /gene="ASPIC1"
     polyA_signal    2642. .2647
                     /gene="ASPIC1"
ORIGIN

  Query Match             86.9%;  Score 1891;  DB 9;  Length 2681;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1941;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         28 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 87

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 147

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 207

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 267

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 327

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 387

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 447

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        448 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCT 507

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        508 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 567

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        568 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 627

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        628 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 687

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        688 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 747

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        748 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 807

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        808 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 867

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        868 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 927

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 987

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1047

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1048 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1107

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1167

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1227

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1228 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1287

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1347

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1407

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1467

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1527

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1587
```

```
Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1647

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1648 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1707

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1708 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1767

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1768 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1827

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1828 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1887

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1888 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1947

Qy       1921 AGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||
Db       1948 AGGATGGCACAGCCTGCGTGGG 1969
```

RESULT 4
AX357100
LOCUS AX357100 2589 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 1 from Patent WO0206478.
ACCESSION AX357100
VERSION AX357100.1 GI:18674281
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
TITLE Extracellular matrix protein
JOURNAL Patent: WO 0206478-A 1 24-JAN-2002;
Akzo Nobel N.V. (NL)
FEATURES Location/Qualifiers
source 1. .2589
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 80.8%; Score 1759; DB 6; Length 2589;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1809; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGGG 1942
             ||||||||||
Db       1845 CCTGCGTGGG 1854


RESULT 5
AX357115
LOCUS       AX357115               2039 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 16 from Patent WO0206478.
ACCESSION   AX357115
VERSION     AX357115.1  GI:18674295
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 16 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2039
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             80.8%;  Score 1758;  DB 6;  Length 2039;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304
```

```
Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
             |||||||||
Db       1845 CCTGCGTGG 1853
```

```
RESULT 6
AX357117
LOCUS       AX357117               2145 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 18 from Patent WO0206478.
ACCESSION   AX357117
VERSION     AX357117.1  GI:18674297
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 18 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2145
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             80.8%;  Score 1758;  DB 6;  Length 2145;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
             |||||||||
Db       1845 CCTGCGTGG 1853


RESULT 7
AX470036
LOCUS       AX470036               1959 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 2 from Patent WO02053709.
ACCESSION   AX470036
VERSION     AX470036.1  GI:22205283
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Richter,W. and Steck,E.
  TITLE     Cartilage marker gene, method for the production and use thereof
  JOURNAL   Patent: WO 02053709-A 2 11-JUL-2002;
            CYTONET GMBH & CO KG (DE)
FEATURES             Location/Qualifiers
     source          1. .1959
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             80.2%;  Score 1745;  DB 6;  Length 1959;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080
```

```
Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

```
RESULT 8
AX470035
LOCUS       AX470035                2507 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 1 from Patent WO02053709.
ACCESSION   AX470035
VERSION     AX470035.1  GI:22205282
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Richter,W. and Steck,E.
  TITLE     Cartilage marker gene, method for the production and use thereof
  JOURNAL   Patent: WO 02053709-A 1 11-JUL-2002;
            CYTONET GMBH & CO KG (DE)
FEATURES             Location/Qualifiers
     source          1. .2507
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             80.2%;  Score 1745;  DB 6;  Length 2507;
```

```
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

```
RESULT 9
HSA279016
LOCUS       HSA279016              2507 bp    mRNA    linear   PRI 11-MAR-2001
DEFINITION  Homo sapiens mRNA for chondrocyte expressed protein 68 kDa (CEP-68
            gene).
ACCESSION   AJ279016
VERSION     AJ279016.1  GI:10178882
KEYWORDS    CEP-68 gene; chondrocyte expressed protein 68 kDa.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Steck,E., Benz,K., Lorenz,H., Loew,M., Gress,T. and Richter,W.
  TITLE     Chondrocyte expressed protein-68 (CEP-68), a novel human marker
            gene for cultured chondrocytes
  JOURNAL   Biochem. J. 353 (Pt 2), 169-174 (2001)
  MEDLINE   21066088
   PUBMED   11139377
REFERENCE   2  (bases 1 to 2507)
  AUTHORS   Steck,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Steck E., Research, Department of
            Orthopaedic Surgery, University of Heidelberg, Schlierbacher
            Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2507
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     gene            1. .2507
                     /gene="CEP-68"
     CDS             1. .1962
                     /gene="CEP-68"
                     /function="unknown"
                     /note="protein with EGF-like calcium binding motif"
                     /codon_start=1
                     /product="chondrocyte expressed protein 68 kDa"
                     /protein_id="CAC08451.1"
                     /db_xref="GI:10178883"
                     /db_xref="GOA:Q9NQ79"
                     /db_xref="SPTREMBL:Q9NQ79"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAA
                     AGAATAAPVLVDGDLNLGSVVKESCEPSC"
     polyA_signal    2469. .2474
                     /gene="CEP-68"
                     /evidence=experimental
ORIGIN
```

```
  Query Match             80.2%;  Score 1745;  DB 9;  Length 2507;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480
Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540
Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600
Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660
Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720
Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780
Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840
Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900
Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960
Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020
Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080
Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140
Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200
Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260
Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320
Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380
Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500
Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560
```

```
Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620
Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680
Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740
Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 10
AX470056
LOCUS AX470056 1947 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 22 from Patent WO02053709.
ACCESSION AX470056
VERSION AX470056.1 GI:22205301
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Richter,W. and Steck,E.
TITLE Cartilage marker gene, method for the production and use thereof
JOURNAL Patent: WO 02053709-A 22 11-JUL-2002;
CYTONET GMBH & CO KG (DE)
FEATURES Location/Qualifiers
source 1. .1947
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 80.1%; Score 1744; DB 6; Length 1947;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60
Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120
Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180
Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240
Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300
Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360
Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420
```

```
Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

```
RESULT 11
HSA421515
LOCUS       HSA421515               1947 bp    mRNA    linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens mRNA for CRTAC1-B protein (CRTAC1 gene).
ACCESSION   AJ421515
VERSION     AJ421515.1  GI:19171210
KEYWORDS    CRTAC1 gene; CRTAC1-B protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Steck,E. and Richter,W.
  TITLE     A novel form of CRTAC1 expressed in human and mouse brain generated
            by alternative splicing of a newly identified last exon
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1947)
  AUTHORS   Steck,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Steck E., Research, Department of
            Orthopaedic Surgery, University of Heidelberg, Schlierbacher
            Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1947
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .1947
                     /gene="CRTAC1"
     CDS             1. .1914
                     /gene="CRTAC1"
                     /codon_start=1
                     /evidence=experimental
                     /product="CRTAC1-B protein"
                     /protein_id="CAD13394.1"
                     /db_xref="GI:19171211"
                     /db_xref="GOA:Q8TE52"
                     /db_xref="SPTREMBL:Q8TE52"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVAQVAFLGGYSSAASRISEPLSRASYL
                     SLGLGLCLQLYAL"
     polyA_signal    1923. .1928
```

/gene="CRTAC1"
ORIGIN

Query Match 80.1%; Score 1744; DB 9; Length 1947;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

RESULT 12
AK057190
LOCUS AK057190 2889 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ32628 fis, clone SYNOV1000118, highly similar to Homo sapiens mRNA for ASPIC (acidic secreted protein in cartilage) (ASPIC1 gene).
ACCESSION AK057190
VERSION AK057190.1 GI:16552786
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T., Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R., Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H., Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,

```
            Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
            Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
            Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SYNOV1000118"
                     /tissue_type="synovial membrane tissue from rheumatioid
                     arthritis"
                     /clone_lib="SYNOV1"
                     /note="cloning vector: pME18SFL3"
ORIGIN

  Query Match             77.5%;  Score 1687;  DB 9;  Length 2889;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1937;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 294

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 354

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 414

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 474

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 534

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 594

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 654

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        655 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCT 714

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 774

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 834

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 894

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 954

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 1014

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 1074

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 GCCGAGGCGTCAGTGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 1134

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 1194

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1195 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1254

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1314

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1315 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCTAGTTCT 1374

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1434

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1494

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1554

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1614

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1674

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1734

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1794

Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1795 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1854

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||
Db       1855 CCAGCAGTGTGGAGGTGACGTGGTCAGATGGCAAGATGGTGAGCCGGAACGCGGCCAGCG 1914

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1974

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 2034

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 2094

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 2154

Qy       1921 AGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||
Db       2155 AGGATGGCACAGCCTGCGTGGG 2176


RESULT 13
AX427468
LOCUS       AX427468               2413 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 41 from Patent WO0123558.
ACCESSION   AX427468
VERSION     AX427468.1  GI:21537614
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hodgson,D.M., Lincoln,S.E., Russo,F.D., Spiro,P.A., Banville,S.C.,
            Bratcher, Shawn,R., Dufour,G.F., Cohen,H.J., Rosen,B.H., Shah,P.,
            Chalup,M.S., Hillman,J.L., Jones,A.L., Yu,J.Y., Greenawalt,L.B.,
            Panzer,S.R., Roseberry,A.M., Wright,R.J., Chen,W., Liu,T.,
            Yap,P.E., Stockdreher,T.K., Amshey,S. and Fong,W.T.
  TITLE     Human secretory molecules
  JOURNAL   Patent: WO 0123558-A 41 05-APR-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2413
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 481454.4.dec"
ORIGIN

  Query Match             74.3%;  Score 1617;  DB 6;  Length 2413;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1917;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 377

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 437

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 497

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 557

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 617

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 677

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 737

Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        738 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 797

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 857

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 917

Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        918 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGATGCTCTGTGGCCTGTGTGGACAGAA 977

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 1037

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 1097

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 1157

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1217

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1277

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1337

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1397

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1457

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1517

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1577

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1637
```

```
Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1697

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1757

Qy       1459 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1817

Qy       1519 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1877

Qy       1579 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1638
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1937

Qy       1639 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1997

Qy       1699 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 1758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 2057

Qy       1759 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 1818
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 2117

Qy       1819 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 1878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 2177

Qy       1879 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 1938
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 2237

Qy       1939 TGG 1941
             |||
Db       2238 TGG 2240
```

RESULT 14
BC034245

```
LOCUS       BC034245               2147 bp    mRNA    linear   PRI 03-OCT-2003
DEFINITION  Homo sapiens cartilage acidic protein 1, mRNA (cDNA clone MGC:16989
            IMAGE:4178363), complete cds.
ACCESSION   BC034245
VERSION     BC034245.1  GI:21706427
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2147)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2147)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 12 Row: h Column: 5
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8922351.
FEATURES             Location/Qualifiers
     source          1. .2147
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:16989 IMAGE:4178363"
                     /tissue_type="Brain, anaplastic oligodendroglioma with
                     1p/19q loss"
                     /clone_lib="NCI_CGAP_Brn67"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .2147
                     /gene="CRTAC1"
                     /note="synonyms: FLJ10320, CEP-68, ASPIC1"
                     /db_xref="LocusID:55118"
                     /db_xref="MIM:606276"
     CDS             249. .1604
                     /codon_start=1
                     /product="CRTAC1 protein"
                     /protein_id="AAH34245.1"
                     /db_xref="GI:21706428"
                     /db_xref="LocusID:55118"
                     /translation="MDPEASDLSRGILALRDVAAEAGVSKYTGGRGVSVGPILSSSAS
                     DIFCDNENGPNFLFHNRGDGTFVDAAASAGVDDPHQHGRGVALADFNRDGKVDIVYGN
                     WNGPHRLYLQMSTHGKVRFRDIASPKFSMPSPVRTVITADFDNDQELEIFFNNIAYRS
                     SSANRLFRVIRREHGDPLIEELNPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESM
                     AQPLSVFRGNQGFNNNWLRVVPRTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCE
                     MEPVAHFGLGKDEASSVEVTWPDGKMVSRNVASGEMNSVLEILYPRDEDTLQDPAPLE
                     CGQGFSQQENGHCMDTNECIQFPFVCPRDKPVCVNTYGSYRCRTNKKCSRGYEPNEDG
                     TACVGTLGQSPGPRPTTPTAAAATAAAAAAAGAATAAPVLVDGDLNLGSVVKESCEPS
                     C"
     misc_feature    1293. .1436
                     /note="EGF_CA; Region: Calcium-binding EGF-like domain"
                     /db_xref="CDD:smart00179"
ORIGIN

  Query Match           66.1%;  Score 1438;  DB 9;  Length 2147;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 60

Qy        565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 120

Qy        625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 180

Qy        685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 240

Qy        745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 300

Qy        805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 360

Qy        865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 420

Qy        925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 480

Qy        985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 540

Qy       1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 600

Qy       1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 660

Qy       1165 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 1224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 720

Qy       1225 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 1284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 780

Qy       1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 840

Qy       1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 900

Qy       1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 960

Qy       1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1020

Qy       1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1080

Qy       1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1140

Qy       1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1200

Qy       1705 TCCTCTACCCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCCTCTACCCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1260

Qy       1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1320

Qy       1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1380

Qy       1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1438


RESULT 15
AX202100
LOCUS       AX202100               2263 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 30 from Patent WO0153531.
ACCESSION   AX202100
VERSION     AX202100.1  GI:15391887
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Phippard,D., Vasanthakamur,G., Dotson,S. and Ma,X.J.
  TITLE     Osteoarthritis tissue derived nucleic acids, polypeptides, vectors,
            and cells
  JOURNAL   Patent: WO 0153531-A 30 26-JUL-2001;
            Pharmacia Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2263
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match            65.6%;  Score 1428;  DB 6;  Length 2263;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 212

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 272

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 332

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 392

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 452

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 512

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||||
Db        513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 572

Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 632

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 692

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 752

Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 812

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 872

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 932

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 992

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1052

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1112

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1172

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1232

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1292

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1352

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1412

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1472

Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1532

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1592

Qy       1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1517
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652

Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712

Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772

Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832

Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCCACTGGAGTGT 1757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCCACTGGAGTGT 1892

Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952

Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012

Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072

Qy       1937 CGTGG 1941
              |||||
Db       2073 CGTGG 2077
```

Search completed: August 10, 2004, 02:06:03
Job time : 8294 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 21:20:25 ; Search time 821 Seconds
(without alignments)
11264.710 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2177	100.0	2177	3	AAA75118	Aaa75118 cDNA enco
2	1994	91.6	2089	6	AAD29792	Aad29792 Human SCI
3	1759	80.8	2589	6	AAD29777	Aad29777 Human SCI
4	1758	80.8	2039	6	AAD29791	Aad29791 Human SCI
5	1758	80.8	2145	6	AAD29793	Aad29793 Human SCI
6	1745	80.2	1959	6	AAL47918	Aal47918 Human cho
7	1745	80.2	2507	6	AAL47917	Aal47917 Human cho
8	1745	80.2	2507	7	ACC50982	Acc50982 Human bla
9	1745	80.2	2507	7	ABX76365	Abx76365 Lung canc
10	1744	80.1	1947	6	AAL47936	Aal47936 Human cho
11	1617	74.3	2413	4	AAS01551	Aas01551 Human sec
12	1565	71.9	1849	6	AAS62301	Aas62301 cDNA sequ
13	1428	65.6	2263	4	AAH23100	Aah23100 Osteoarth
14	1349	62.0	1574	4	AAI60925	Aai60925 Human pol
15	1314	60.4	1501	4	AAI59139	Aai59139 Human pol
16	1314	60.4	1501	8	ADB49122	Adb49122 Novel hum
17	1236	56.8	2178	4	AAH14000	Aah14000 Human cDN
18	1236	56.8	2178	7	ACC50981	Acc50981 Human bla
19	1236	56.8	2178	7	ABX76364	Abx76364 Lung canc
20	687	31.6	4794	7	ACC50983	Acc50983 Human bla
21	687	31.6	4794	7	ABX76366	Abx76366 Lung canc
22	649	29.8	883	6	ABL42156	Abl42156 Nucleotid
23	648	29.8	1143	3	AAZ65257	Aaz65257 Human sec
24	648	29.8	1143	7	ADA56306	Ada56306 Gene enco
25	648	29.8	1143	7	ADA40142	Ada40142 Human sec
26	648	29.8	1143	9	ADE11646	Ade11646 Human sec
27	496	22.8	608	4	AAH05239	Aah05239 Human cDN
28	419	19.2	465	8	ACH31964	Ach31964 Human end
29	358	16.4	447	8	ACH42110	Ach42110 Human foe
c 30	191	8.8	1400	5	AAS83195	Aas83195 DNA encod
31	69	3.2	319	2	AAT25375	Aat25375 Human gen
32	60	2.8	60	6	ABN43600	Abn43600 Human spl
33	57	2.6	57	6	AAL47919	Aal47919 Human cho
34	41	1.9	41	6	ABL42162	Abl42162 Probe #1
35	41	1.9	41	6	ABL42161	Abl42161 Probe #2
36	40	1.8	752	4	AAH23149	Aah23149 Osteoarth
c 37	25	1.1	25	6	AAD29782	Aad29782 5' RACE p
38	25	1.1	33	6	ABL42159	Abl42159 PCR prime
39	24	1.1	24	6	ABL42157	Abl42157 PCR prime
40	23	1.1	23	6	AAD29784	Aad29784 3' RACE p
41	22	1.0	22	6	AAD29785	Aad29785 Primer #2
c 42	22	1.0	22	6	AAD29783	Aad29783 Primer #1
43	21	1.0	21	6	AAL47931	Aal47931 Human cho
c 44	21	1.0	21	6	AAD29779	Aad29779 Human SCI
c 45	21	1.0	1216	6	ABQ28694	Abq28694 Oligonucl

ALIGNMENTS

RESULT 1
AAA75118
ID AAA75118 standard; cDNA; 2177 BP.
XX
AC AAA75118;
XX
DT 15-JAN-2001 (first entry)
XX
DE cDNA encoding a human secretory protein.
XX
KW Human; secretory protein; HSECP; cancer; gastrointestinal disorder;
KW inflammation; cardiovascular disorder; neurological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 123. .2003
FT /*tag= a
FT /product= "secretory protein"
FT sig_peptide 123. .203
FT /*tag= b
XX
PN WO200052151-A2.
XX
PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-US005621.
XX
PR 05-MAR-1999; 99US-0123117P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Baughn MR, Yue H, Au-Young J, Lu DAM, Azimzai Y;
XX
DR WPI; 2000-579282/54.
DR P-PSDB; AAB08864.
XX
PT Twenty two human secretory proteins for diagnosing, treating and
PT preventing cancer, inflammation, and gastrointestinal, cardiovascular and
PT neurological disorders.
XX
PS Claim 4; Page 102-103; 107pp; English.
XX
CC The present sequence encodes a human secretory protein, designated HSECP-
CC 1. The specification also describes HSECP-2 to HSECP-22. The proteins are
CC useful for diagnosing, treating and preventing cancer, inflammation, and

CC gastrointestinal, cardiovascular and neurological disorders. The proteins
CC may also be used to identify agonists, antagonists, and inhibitors. The
CC polynucleotides may be used for producing the protein recombinantly, and
CC as a source of probes and primers for isolating and identifying related
CC sequences
XX
SQ Sequence 2177 BP; 466 A; 627 C; 660 G; 424 T; 0 U; 0 Other;

Query Match 100.0%; Score 2177; DB 3; Length 2177;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCA 360

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560

Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Qy       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980

Qy       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040

Qy       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100

Qy       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160

Qy       2161 TCTTGAAAAAAAAAAAA 2177
              |||||||||||||||||
Db       2161 TCTTGAAAAAAAAAAAA 2177
```

RESULT 2
AAD29792
ID AAD29792 standard; cDNA; 2089 BP.
XX
AC AAD29792;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 splice variant B cDNA.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   59. .1915
FT   /*tag= a
FT   /product= "Human SCIM-1 splice variant B protein"
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR P-PSDB; AAE18681.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 4; Page 36-37; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for
CC induction of immunological tolerance to an autoantigen in patients
CC suffering from autoimmune disorders, more specifically rheumatoid
CC arthritis. The inflammatory disorders treated include diseases like
CC Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC multiple sclerosis and diabetes. Polypeptides of the invention are also
CC useful in a diagnostic method for the detection of activated autoreactive
CC T cells. The present sequence is human SCIM-1 splice variant B cDNA
XX
SQ Sequence 2089 BP; 462 A; 592 C; 614 G; 421 T; 0 U; 0 Other;

Query Match 91.6%; Score 1994; DB 6; Length 2089;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1992
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1845 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1904

Qy       1993 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 2052
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1905 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 1964

Qy       2053 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2112
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1965 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2024

Qy       2113 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2025 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2084

Qy       2173 AAAAA 2177
             |||||
Db       2085 AAAAA 2089
```

```
RESULT 3
AAD29777
ID   AAD29777 standard; cDNA; 2589 BP.
XX
AC   AAD29777;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 protein encoding cDNA.
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             59. .2020
FT                   /*tag= a
FT                   /product= "Human SCIM-1 protein"
FT   sig_peptide     59. .121
FT                   /*tag= d
FT   sig_peptide     59. .115
FT                   /*tag= b
FT   mat_peptide     116. .2017
FT                   /*tag= c
FT                   /product= "Human SCIM-1 mature protein #1"
FT   mat_peptide     122. .2017
FT                   /*tag= e
FT                   /product= "Human SCIM-1 mature protein #2"
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   P-PSDB; AAE18679.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 4; Page 29-30; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
```

```
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 protein encoding cDNA
XX
SQ   Sequence 2589 BP; 585 A; 726 C; 765 G; 512 T; 0 U; 1 Other;

  Query Match            80.8%;  Score 1759;  DB 6;  Length 2589;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy       193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy       253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy       313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy       373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy       433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy       493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy       553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy       613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy       673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy       733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy       793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy       853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy       913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy       973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy      1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy      1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy      1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy      1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy      1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy      1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy      1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy      1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy      1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy      1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy      1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy      1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy      1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy      1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy      1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy      1933 CCTGCGTGGG 1942
            ||||||||||
Db      1845 CCTGCGTGGG 1854


RESULT 4
AAD29791
ID   AAD29791 standard; cDNA; 2039 BP.
XX
AC   AAD29791;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 splice variant A cDNA.
```

```
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             59. .1972
FT                   /*tag=  a
FT                   /product= "Human SCIM-1 splice variant A protein"
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   P-PSDB; AAE18680.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 4; Page 35-36; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 splice variant A cDNA
XX
SQ   Sequence 2039 BP; 448 A; 586 C; 596 G; 409 T; 0 U; 0 Other;

  Query Match             80.8%;  Score 1758;  DB 6;  Length 2039;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
              |||||||||
Db       1845 CCTGCGTGG 1853


RESULT 5
AAD29793
ID   AAD29793 standard; cDNA; 2145 BP.
XX
AC   AAD29793;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 splice variant C cDNA.
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             59. .1897
FT                   /*tag=  a
FT                   /product= "Human SCIM-1 splice variant C protein"
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   P-PSDB; AAE18682.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 4; Page 37; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 splice variant C cDNA
XX
SQ   Sequence 2145 BP; 497 A; 604 C; 611 G; 433 T; 0 U; 0 Other;

  Query Match           80.8%;  Score 1758;  DB 6;  Length 2145;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
              |||||||||
Db       1845 CCTGCGTGG 1853
```

RESULT 6
AAL47918
ID AAL47918 standard; cDNA; 1959 BP.
XX
AC AAL47918;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:2.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein; gene; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .1959
FT		/*tag= a
FT		/product= "CEP-68"
FT		/partial
FT		/note= "no stop codon"
FT	sig_peptide	1. .57
FT		/*tag= b
FT	mat_peptide	58. .1959
FT		/*tag= c

XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR P-PSDB; AAO18264.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 2; Page 92; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is a coding sequence of the CEP-68 protein
XX
SQ Sequence 1959 BP; 394 A; 590 C; 596 G; 379 T; 0 U; 0 Other;

Query Match 80.2%; Score 1745; DB 6; Length 1959;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db       181 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       961 GTCCGCTTCCGGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

```
RESULT 7
AAL47917
ID   AAL47917 standard; cDNA; 2507 BP.
XX
AC   AAL47917;
XX
DT   26-SEP-2002  (first entry)
XX
DE   Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:1.
XX
KW   Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW   extracellular matrix protein; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1962
FT                   /*tag=  a
FT                   /product= "CEP-68"
FT   sig_peptide     1. .57
FT                   /*tag=  b
FT   mat_peptide     58. .1959
FT                   /*tag=  c
XX
PN   WO200253709-A1.
```

XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR P-PSDB; AA018264.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 1; Page 91-92; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is a coding sequence of the CEP-68 protein
XX
SQ Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

Query Match 80.2%; Score 1745; DB 6; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796


RESULT 8
ACC50982
ID   ACC50982 standard; cDNA; 2507 BP.
XX
AC   ACC50982;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Human bladder cancer associated cDNA sequence SEQ ID NO:57.
XX
KW   Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003003906-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US021338.
XX
PR   03-JUL-2001; 2001US-0302814P.
PR   03-AUG-2001; 2001US-0310099P.
PR   08-NOV-2001; 2001US-0343705P.
PR   13-NOV-2001; 2001US-0350666P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack DH,  Aziz N;
XX
DR   WPI; 2003-201532/19.
DR   P-PSDB; ABR48171.
XX
PT   Detecting a bladder cancer-associated transcript in a cell from a
PT   patient, comprises contacting a biological sample from the patient with a
PT   bladder cancer-associated polynucleotide or antibody.
XX
PS   Claim 6; Page 248-249; 307pp; English.
XX
CC   The present invention describes a method for detecting a bladder cancer-
CC   associated transcript in a cell from a patient. The method comprises
CC   contacting a biological sample from the patient with a polynucleotide
CC   that selectively hybridises to a sequence that is 80 % identical to a
CC   table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC   encode the human bladder cancer-associated proteins given in ABR48146 to
CC   ABR48242). Bladder cancer-associated sequences from the present invention
CC   have cytostatic activities, and can be used in antisense gene therapy and
CC   in vaccine production. The method can be used for detecting a bladder
CC   cancer-associated transcript in a cell from a patient. The method is
CC   useful in diagnosing or treating bladder cancer and in screening for
CC   compounds that modulate bladder cancer, such as hormones or antibodies.
CC   The nucleic acid molecules from the present invention may be used in
CC   various screening and diagnostic methods, and for gene therapy, vaccine
CC   and/or antisense/inhibition applications
XX
SQ   Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

  Query Match            80.2%;  Score 1745;  DB 7;  Length 2507;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796


RESULT 9
ABX76365
ID   ABX76365 standard; DNA; 2507 BP.
XX
AC   ABX76365;
XX
DT   02-APR-2003  (first entry)
XX
DE   Lung cancer-associated polynucleotide #229.
XX
KW   Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
KW   antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW   small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW   chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW   interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS   Unidentified.
XX
PN   WO200286443-A2.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US012476.
XX
PR   18-APR-2001; 2001US-0284770P.
PR   10-MAY-2001; 2001US-0290492P.
PR   09-NOV-2001; 2001US-0339245P.
PR   13-NOV-2001; 2001US-0350666P.
PR   29-NOV-2001; 2001US-0334370P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Aziz N,  Murray R;
XX
DR   WPI; 2003-093161/08.
DR   P-PSDB; ABU56636.
XX
PT   Detecting a lung cancer-associated transcript in a cell from a patient
PT   for treating lung cancer, by contacting a biological sample from the
PT   patient with a polynucleotide that exhibits increased or decreased
PT   expression in lung cancer.
XX
PS   Claim 22; Page 366; 453pp; English.
XX
CC   The invention relates to a method for detecting a lung cancer-associated
CC   transcript in a cell from a patient, comprising contacting a biological
CC   sample from the patient with a polynucleotide that selectively hybridises
CC   to a sequence that is at least 80 % identical to a gene that exhibits
CC   increased or decreased expression in lung cancer samples. Lung cancer-
CC   associated polynucleotides and polypeptides are used for identifying a
CC   compound that modulates a lung cancer-associated polypeptide, for
CC   inhibiting proliferation of a lung cancer-associated cell to treat lung
CC   cancer in a patient and for treating a mammal having lung cancer by
CC   administering a modulatory compound identified. The methods are useful
CC   for treating lung cancer, such as small cell lung cancer, non-small cell
CC   lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC   emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC   hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC   bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC   for diagnostic purposes and as targets for screening for therapeutic
CC   compounds that modulate lung cancer, such as antibodies. Sequences
CC   ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
CC   invention
XX
SQ   Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

  Query Match            80.2%;  Score 1745;  DB 7;  Length 2507;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480
```

```
Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 10
AAL47936
ID AAL47936 standard; cDNA; 1947 BP.
XX
AC AAL47936;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:22.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein; gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1. .1914
FT   /*tag= a
FT   /product= "alternatively spliced CEP-68"
FT   /transl_except= (pos:1129. .1131,aa:Glu)
FT   /transl_except= (pos:1297. .1299,aa:Phe)
FT   /transl_except= (pos:1621. .1623,aa:Val)
FT   /transl_except= (pos:1648. .1650,aa:Asn)
FT   /transl_except= (pos:1669. .1674,aa:Phe)
XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR P-PSDB; AAO18267.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 4; Page 97-98; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is an alternatively spliced coding sequence of the CEP-
CC 68 protein
XX
SQ Sequence 1947 BP; 413 A; 566 C; 571 G; 397 T; 0 U; 0 Other;

  Query Match   80.1%; Score 1744; DB 6; Length 1947;
  Best Local Similarity 99.9%; Pred. No. 0;

```
  Matches 1794;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

```
RESULT 11
AAS01551
ID   AAS01551 standard; cDNA; 2413 BP.
XX
AC   AAS01551;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secretory molecule cDNA sptm #41.
XX
KW   Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW   cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW   immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW   nervous system disease; mental retardation; developmental disorder;
KW   neuromuscular disorder; microarray; Incyte ID number 4814544dec; ss.
XX
OS   Homo sapiens.
XX
PN   WO200123558-A2.
XX
PD   05-APR-2001.
XX
PF   19-SEP-2000; 2000WO-US025610.
```

XX
PR 28-SEP-1999; 99US-0156624P.
PR 28-SEP-1999; 99US-0156625P.
PR 02-DEC-1999; 99US-0168611P.
PR 02-DEC-1999; 99US-0168613P.
PR 02-DEC-1999; 99US-0168614P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM;
PI Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK, Amshey S;
PI Fong WT;
XX
DR WPI; 2001-258134/26.
XX
PT New secretory polynucleotides (SPTM) and the polypeptides they encode,
PT for use in inducing antibodies and screening libraries of compounds.
XX
PS Claim 1; Page 146; 161pp; English.
XX
CC The present sequence for human secretory molecule cDNA sptm #41 (Incyte
CC ID number 4814544dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-
CC AAS01573) which encode for the secretory polypeptides SPTM. The sptm
CC polynucleotides are useful for screening a compound for effectiveness in
CC altering expression of a target polynucleotide, where the target
CC polynucleotide comprises sptm. Sptm is also useful in a method for
CC assessing the toxicity of a test compound. Sptm and its fragments or
CC complementary sequences, may be used to identify the presence of and/or
CC determine the degree of similarity between two nucleic acid sequences.
CC Sptm can also be used for a variety of diagnostic and therapeutic
CC purposes, e.g. diagnosing a particular condition, disease or disorder
CC associated with cell signalling, such conditions include cell
CC proliferative disorders such as atherosclerosis, and cancers including
CC leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC disease, nutritional and metabolic disease of the nervous system, mental
CC retardation and other developmental disorders, and muscular dystrophy and
CC other neuromuscular disorders. Sptm can also be used to design probes
CC useful in diagnostic assays, which may be used to monitor the progress of
CC conditions or disorders associated with abnormal levels of expression of
CC sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC gene therapy, for inducing antibodies, or in microarrays
XX
SQ Sequence 2413 BP; 511 A; 714 C; 722 G; 459 T; 0 U; 7 Other;

Query Match 74.3%; Score 1617; DB 4; Length 2413;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1917; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy        19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       318 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 377

Qy        79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       378 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 437

Qy       139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       438 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 497

Qy       199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       498 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 557

Qy       259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       558 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 617

Qy       319 ATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       618 ATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 677

Qy       379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       678 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 737

Qy       439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       738 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 797

Qy       499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       798 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 857

Qy       559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       858 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 917

Qy       619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       918 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGATGCTCTGTGGCCTGTGTGGACAGAA 977

Qy       679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       978 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 1037

Qy       739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1038 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 1097

Qy       799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1098 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 1157

Qy       859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1158 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1217

Qy       919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1218 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1277

Qy       979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1278 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1337

Qy      1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1338 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1397

Qy      1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1398 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1457

Qy      1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1458 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1517

Qy      1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1518 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1577

Qy      1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1578 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1637

Qy      1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1638 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1697

Qy      1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1698 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1757
```

```
Qy       1459 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1817

Qy       1519 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1877

Qy       1579 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1638
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1937

Qy       1639 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1997

Qy       1699 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 1758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 2057

Qy       1759 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 1818
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 2117

Qy       1819 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 1878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 2177

Qy       1879 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 1938
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 2237

Qy       1939 TGG 1941
             |||
Db       2238 TGG 2240
```

```
RESULT 12
AAS62301
ID   AAS62301 standard; cDNA; 1849 BP.
XX
AC   AAS62301;
XX
DT   14-FEB-2002  (first entry)
XX
DE   cDNA sequence #88 encoding novel human secreted protein.
XX
KW   Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW   immune deficiency disorder; blood disorder; inflammatory disorder;
KW   infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW   immunosuppressive; antirheumatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200177291-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US010485.
XX
PR   06-APR-2000; 2000US-0195604P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Wong GG,  Clark HF,  Fechtel K,  Agostino MJ,  Howes SH,  Resnick RJ;
PI   Gulukota K,  Graham JR;
XX
DR   WPI; 2002-010900/01.
XX
PT   New polynucleotides encoding secreted proteins useful for treating e.g.
PT   asthma, HIV and Crohn's disease.
XX
PS   Claim 1; Page 122; 391pp; English.
XX
CC   The present invention relates to the isolation of novel cDNA sequences
CC   which encode human secreted proteins. The cDNA sequences have been
CC   derived from a variety of human tissues. The invention also provides a
CC   method for producing proteins from these polynucleotide sequences. The
CC   proteins are useful for identifying compounds that modulate their
CC   activity and production, and the cell is also useful for identifying
CC   compounds that modulate expression of the polynucleotide sequences
CC   encoding the secreted proteins. The sequences of the invention are useful
CC   for treating diseases such as hyperproliferative disorders (e.g. cancer),
CC   immune deficiency disorders (e.g. severe combined immunodeficiency
CC   (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
CC   (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and
CC   infectious disorders (e.g. hepatitis). The polynucleotide sequences of
CC   the invention are also useful in gene therapy. AAS62214-AAS62838
CC   represent the cDNA sequences of the invention that encode for novel human
CC   secreted proteins
XX
SQ   Sequence 1849 BP; 395 A; 535 C; 544 G; 372 T; 0 U; 3 Other;
```

```
  Query Match           71.9%;  Score 1565;  DB 6;  Length 1849;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1665;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 60

Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 120

Qy        359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 180

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        181 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGC 240

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 300

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 360

Qy        599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 420

Qy        659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 480

Qy        719 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 540

Qy        779 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 600

Qy        839 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 660

Qy        899 CAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 958
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGAGAATGGGCCTAAATTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 720
```

```
Qy        959 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 780

Qy       1019 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 840

Qy       1079 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 900

Qy       1139 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 960

Qy       1199 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1020

Qy       1259 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1080

Qy       1319 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1140

Qy       1379 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1200

Qy       1439 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1260

Qy       1499 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1320

Qy       1559 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1380

Qy       1619 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1440

Qy       1679 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1500

Qy       1739 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1560

Qy       1799 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1620

Qy       1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGG 1905
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGG 1667
```

```
RESULT 13
AAH23100
ID   AAH23100 standard; DNA; 2263 BP.
XX
AC   AAH23100;
XX
DT   17-SEP-2001  (first entry)
XX
DE   Osteoarthritis tissue-derived nucleic acid sequence #30.
XX
KW   Osteoarthritis; infectious disorder; Crohn's disease; sepsis; human;
KW   wound healing; osteopathic; anti-arthritic, anti-inflammatory; vulnerary;
KW   antibacterial; antiallergic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200153531-A2.
XX
PD   26-JUL-2001.
XX
PF   18-JAN-2001; 2001WO-US000016.
XX
PR   18-JAN-2000; 2000US-0176523P.
XX
PA   (PHAA ) PHARMACIA CORP.
XX
PI   Phippard D,  Vasanthakamur G,  Dotson S,  Ma X;
XX
DR   WPI; 2001-451914/48.
XX
PT   Substantially purified protein, polypeptide or their fragments, used to
PT   identify a biologically active compound or composition and treat
PT   mammalian osteoarthritis.
XX
PS   Claim 1; Page 115-116; 144pp; English.
XX
CC   Sequences AAH23071-23152 represent nucleic acid sequences derived from
CC   osteoarthritis tissues. The sequences are useful as probes and for the
CC   diagnosis or prognosis of mammalian osteoarthritis. The polynucleotides
CC   and polypeptides of the invention are useful for generating diagnostic
CC   reagents, as targets for small molecule drug development, generation of
CC   therapeutics, and cloning genes. Specific antibodies are used to generate
CC   enzyme linked immunosorbent assays for detection of osteoarthritis. The
CC   invented molecules can be used to treat osteoarthritis or to analyse the
CC   disease-modifying activity of osteoarthritis drugs. Other disorders
CC   treatable using the nucleic acid sequences include atopic, inflammatory
CC   and infectious disorders e.g. Crohn's disease and sepsis, and wound
CC   healing
XX
SQ   Sequence 2263 BP; 485 A; 667 C; 676 G; 428 T; 0 U; 7 Other;

  Query Match           65.6%;  Score 1428;  DB 4;  Length 2263;
  Best Local Similarity  99.6%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 212

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 272

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 332

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 392

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 452

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 512

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
```

```
Db        513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 572
Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
Db        573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 632
Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
Db        633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 692
Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
Db        693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 752
Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
Db        753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 812
Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
Db        813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 872
Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
Db        873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 932
Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
Db        933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 992
Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
Db        993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1052
Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
Db       1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1112
Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
Db       1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1172
Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
Db       1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1232
Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
Db       1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1292
Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
Db       1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1352
Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
Db       1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1412
Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
Db       1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1472
Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
Db       1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1532
Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
Db       1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1592
Qy       1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1517
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652
```

```
Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577
Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712
Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637
Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772
Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697
Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832
Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1757
Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1892
Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816
Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952
Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012
Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936
Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072
Qy       1937 CGTGG 1941
Db       2073 CGTGG 2077
```

```
RESULT 14
AAI60925
ID   AAI60925 standard; cDNA; 1574 BP.
XX
AC   AAI60925;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4914.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J,  Zhao QA;
```

```
PI   Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41769.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.
XX
PS   Claim 1; SEQ ID NO 4914; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification
XX
SQ   Sequence 1574 BP; 358 A; 445 C; 455 G; 316 T; 0 U; 0 Other;

  Query Match             62.0%;  Score 1349;  DB 4;  Length 1574;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1349;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        593 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGG 652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGG 60

Qy        653 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 120

Qy        713 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 180

Qy        773 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 240

Qy        833 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 300

Qy        893 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 360

Qy        953 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 420

Qy       1013 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 480

Qy       1073 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 540

Qy       1133 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 600

Qy       1193 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 660

Qy       1253 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 720

Qy       1313 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 780

Qy       1373 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 1432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 840

Qy       1433 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 1492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 900

Qy       1493 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 1552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 960

Qy       1553 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1020

Qy       1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1080

Qy       1673 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1140

Qy       1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1200

Qy       1793 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1260

Qy       1853 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1320

Qy       1913 GCCCAACGAGGATGGCACAGCCTGCGTGG 1941
              |||||||||||||||||||||||||||||
Db        1321 GCCCAACGAGGATGGCACAGCCTGCGTGG 1349


RESULT 15
AAI59139
ID   AAI59139 standard; cDNA; 1501 BP.
XX
AC   AAI59139;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1342.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
```

PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39983.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 1342; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1501 BP; 319 A; 436 C; 441 G; 305 T; 0 U; 0 Other;

Query Match 60.4%; Score 1314; DB 4; Length 1501;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        628 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 95

Qy        688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 155

Qy        748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 215

Qy        808 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 275

Qy        868 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 335

Qy        928 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 395

Qy        988 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 455

Qy       1048 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 515

Qy       1108 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 575

Qy       1168 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 635

Qy       1228 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 695

Qy       1288 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGGCACAGGGGGTGTGG 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 755

Qy       1348 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 815

Qy       1408 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        816 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 875

Qy       1468 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 935

Qy       1528 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995

Qy       1588 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        996 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055

Qy       1648 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1115

Qy       1708 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1116 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1175

Qy       1768 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1176 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1235

Qy       1828 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1236 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1295

Qy       1888 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1296 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1349
```

Search completed: August 9, 2004, 23:47:42
Job time : 827 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 23:17:15 ; Search time 178 Seconds
(without alignments)
6787.236 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1314	60.4	1501	4	US-09-620-312D-1032	Sequence 1032, Ap
	2	738	33.9	789	4	US-09-023-655-484	Sequence 484, App
	3	109	5.0	707	4	US-09-023-655-209	Sequence 209, App
	4	20	0.9	411	4	US-09-508-824-24	Sequence 24, Appl
	5	20	0.9	414	4	US-09-508-824-23	Sequence 23, Appl
	6	20	0.9	431	4	US-09-508-824-26	Sequence 26, Appl
	7	20	0.9	434	4	US-09-508-824-25	Sequence 25, Appl
	8	20	0.9	500	4	US-09-508-824-27	Sequence 27, Appl
	9	20	0.9	817	4	US-09-508-824-13	Sequence 13, Appl
	10	20	0.9	953	4	US-09-508-824-17	Sequence 17, Appl
	11	20	0.9	1001	4	US-09-508-824-18	Sequence 18, Appl
	12	20	0.9	1133	4	US-09-508-824-14	Sequence 14, Appl
	13	20	0.9	1134	4	US-09-508-824-20	Sequence 20, Appl
	14	20	0.9	1173	4	US-09-508-824-19	Sequence 19, Appl
	15	20	0.9	1185	4	US-09-508-824-15	Sequence 15, Appl
	16	20	0.9	1186	4	US-09-508-824-21	Sequence 21, Appl
	17	20	0.9	1916	4	US-09-508-824-16	Sequence 16, Appl
	18	20	0.9	1916	4	US-09-508-824-22	Sequence 22, Appl
	19	20	0.9	2332	4	US-09-508-824-2	Sequence 2, Appli
c	20	20	0.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	21	20	0.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	22	19	0.9	299	4	US-09-132-316-11	Sequence 11, Appl
	23	19	0.9	299	4	US-09-132-316-45	Sequence 45, Appl
	24	19	0.9	326	4	US-09-132-316-44	Sequence 44, Appl
	25	19	0.9	403	3	US-09-328-111-360	Sequence 360, App
c	26	19	0.9	422	4	US-09-132-316-25	Sequence 25, Appl
	27	19	0.9	461	4	US-09-132-316-20	Sequence 20, Appl
	28	19	0.9	642	4	US-09-252-991A-11468	Sequence 11468, A
	29	19	0.9	747	4	US-09-252-991A-11489	Sequence 11489, A
	30	19	0.9	768	4	US-09-235-451-6	Sequence 6, Appli
	31	19	0.9	1281	3	US-09-105-537-19	Sequence 19, Appl
	32	19	0.9	1512	4	US-09-199-637A-171	Sequence 171, App
c	33	19	0.9	1600	4	US-09-434-288-10	Sequence 10, Appl
	34	19	0.9	1790	4	US-09-484-970B-67	Sequence 67, Appl
	35	19	0.9	1977	4	US-09-548-372D-5	Sequence 5, Appli
	36	19	0.9	1977	4	US-09-548-367D-5	Sequence 5, Appli
	37	19	0.9	1977	4	US-09-551-853D-5	Sequence 5, Appli
	38	19	0.9	2106	1	US-07-828-790B-1	Sequence 1, Appli
c	39	19	0.9	2184	4	US-09-484-970B-161	Sequence 161, App
	40	19	0.9	2380	4	US-09-235-451-35	Sequence 35, Appl
	41	19	0.9	2779	4	US-09-149-476-191	Sequence 191, App
	42	19	0.9	2805	4	US-09-132-316-1	Sequence 1, Appli
	43	19	0.9	2860	4	US-09-149-476-314	Sequence 314, App
	44	19	0.9	2913	1	US-08-243-542-7	Sequence 7, Appli
	45	19	0.9	2913	1	US-08-477-407-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-620-312D-1032
; Sequence 1032, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1032
; LENGTH: 1501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(1468)
US-09-620-312D-1032

Query Match 60.4%; Score 1314; DB 4; Length 1501;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        628 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 95
```

```
Qy       688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
Db        96 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 155

Qy       748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
Db       156 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 215

Qy       808 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 867
Db       216 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 275

Qy       868 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 927
Db       276 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 335

Qy       928 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 987
Db       336 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 395

Qy       988 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 1047
Db       396 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 455

Qy      1048 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 1107
Db       456 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 515

Qy      1108 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 1167
Db       516 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 575

Qy      1168 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 1227
Db       576 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 635

Qy      1228 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 1287
Db       636 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 695

Qy      1288 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 1347
Db       696 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 755

Qy      1348 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407
Db       756 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 815

Qy      1408 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 1467
Db       816 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 875

Qy      1468 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 1527
Db       876 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 935

Qy      1528 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
Db       936 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995

Qy      1588 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647
Db       996 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055

Qy      1648 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1707
Db      1056 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1115

Qy      1708 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1767
Db      1116 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1175

Qy      1768 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1827
Db      1176 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1235

Qy      1828 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1887
Db      1236 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1295

Qy      1888 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
Db      1296 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1349
```

RESULT 2
US-09-023-655-484
; Sequence 484, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 484:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 192279
US-09-023-655-484

Query Match 33.9%; Score 738; DB 4; Length 789;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1058 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 1117
Db         1 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 60

Qy      1118 GGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 1177
Db        61 GGACATCGNCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 120
```

```
Qy       1178 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 180

Qy       1238 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 240

Qy       1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 300

Qy       1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 360

Qy       1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 420

Qy       1478 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 480

Qy       1538 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 540

Qy       1598 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 1657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 600

Qy       1658 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 660

Qy       1718 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 1777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 720

Qy       1778 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 1837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 780

Qy       1838 AGACAAGCC 1846
             |||||||||
Db        781 AGACAAGCC 789


RESULT 3
US-09-023-655-209
; Sequence 209, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SYNORAB01
;      CLONE:  079378
US-09-023-655-209

  Query Match             5.0%;  Score 109;  DB 4;  Length 707;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-41;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 1893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 60

Qy       1894 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 109


RESULT 4
US-09-508-824-24
; Sequence 24, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 411
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-24

  Query Match             0.9%;  Score 20;  DB 4;  Length 411;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db         15 GGCGGCGGCGCCGGCGGCAG 34


RESULT 5
US-09-508-824-23
```

```
; Sequence 23, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 414
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-23

  Query Match             0.9%;  Score 20;  DB 4;  Length 414;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db         15 GGCGGCGGCGCCGGCGGCAG 34


RESULT 6
US-09-508-824-26
; Sequence 26, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-26

  Query Match             0.9%;  Score 20;  DB 4;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db         15 GGCGGCGGCGCCGGCGGCAG 34


RESULT 7
US-09-508-824-25
; Sequence 25, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 434
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-25

  Query Match             0.9%;  Score 20;  DB 4;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db         15 GGCGGCGGCGCCGGCGGCAG 34


RESULT 8
US-09-508-824-27
; Sequence 27, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
;   LENGTH: 500
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-27

  Query Match             0.9%;  Score 20;  DB 4;  Length 500;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db         81 GGCGGCGGCGCCGGCGGCAG 100


RESULT 9
US-09-508-824-13
; Sequence 13, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
```

```
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 817
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-13

  Query Match             0.9%;  Score 20;  DB 4;  Length 817;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        106 GGCGGCGGCGCCGGCGGCAG 125


RESULT 10
US-09-508-824-17
; Sequence 17, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 953
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-17

  Query Match             0.9%;  Score 20;  DB 4;  Length 953;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        201 GGCGGCGGCGCCGGCGGCAG 220


RESULT 11
US-09-508-824-18
; Sequence 18, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1001
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-18

  Query Match             0.9%;  Score 20;  DB 4;  Length 1001;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        216 GGCGGCGGCGCCGGCGGCAG 235


RESULT 12
US-09-508-824-14
; Sequence 14, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 1133
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-14

  Query Match             0.9%;  Score 20;  DB 4;  Length 1133;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        369 GGCGGCGGCGCCGGCGGCAG 388


RESULT 13
US-09-508-824-20
; Sequence 20, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 1134
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-20

  Query Match             0.9%;  Score 20;  DB 4;  Length 1134;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
```

```
Db        369 GGCGGCGGCGCCGGCGGCAG 388


RESULT 14
US-09-508-824-19
; Sequence 19, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-19

  Query Match             0.9%;  Score 20;  DB 4;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        428 GGCGGCGGCGCCGGCGGCAG 447


RESULT 15
US-09-508-824-15
; Sequence 15, Application US/09508824
; Patent No. 6635811
; GENERAL INFORMATION:
;  APPLICANT: Flintham, John E
;  APPLICANT:  Gale, Michael D
;  APPLICANT:  Holdsworth, Michael J
;  TITLE OF INVENTION: Pre-harvest Sprouting
;  FILE REFERENCE: Mewburn
;  CURRENT APPLICATION NUMBER: US/09/508,824
;  CURRENT FILING DATE:  2000-05-17
;  PRIOR APPLICATION NUMBER: PCT/GB98/02835
;  PRIOR FILING DATE: 1998-09-18
;  PRIOR APPLICATION NUMBER: GB 9720060.4
;  PRIOR FILING DATE: 1997-09-19
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 1185
;   TYPE: DNA
;   ORGANISM: Triticum aestivum
US-09-508-824-15

  Query Match             0.9%;  Score 20;  DB 4;  Length 1185;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         45 GGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||
Db        419 GGCGGCGGCGCCGGCGGCAG 438


Search completed: August 10, 2004, 03:35:34
Job time : 183 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 10, 2004, 02:06:12 ; Search time 5598 Seconds
(without alignments)
1906.777 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3222919 seqs, 2451570024 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1994	91.6	2089	12	US-10-332-929A-17	Sequence 17, Appl
2	1759	80.8	2589	12	US-10-332-929A-1	Sequence 1, Appli
3	1758	80.8	2039	12	US-10-332-929A-16	Sequence 16, Appl
4	1758	80.8	2145	12	US-10-332-929A-18	Sequence 18, Appl
5	1745	80.2	2507	16	US-10-295-027-182	Sequence 182, App
6	1745	80.2	2507	17	US-10-188-832-57	Sequence 57, Appl
7	1565	71.9	1849	9	US-09-822-830A-88	Sequence 88, Appl
8	1428	65.6	2263	9	US-09-765-231A-30	Sequence 30, Appl
9	1314	60.4	1501	15	US-10-037-270-1032	Sequence 1032, Ap
10	1314	60.4	1501	16	US-10-117-722-1032	Sequence 1032, Ap
11	1236	56.8	2178	16	US-10-295-027-180	Sequence 180, App
12	1236	56.8	2178	17	US-10-188-832-55	Sequence 55, Appl
13	738	33.9	789	17	US-10-641-643-484	Sequence 484, App
14	687	31.6	4794	16	US-10-295-027-184	Sequence 184, App
15	687	31.6	4794	17	US-10-188-832-59	Sequence 59, Appl
16	648	29.8	1143	10	US-09-892-877-18	Sequence 18, Appl
17	648	29.8	1143	10	US-09-948-783-18	Sequence 18, Appl
18	419	19.2	465	10	US-09-918-995-19176	Sequence 19176, A
19	358	16.4	447	10	US-09-918-995-29322	Sequence 29322, A
20	225	10.3	399	13	US-10-085-783A-21919	Sequence 21919, A
21	225	10.3	399	16	US-10-242-535A-21919	Sequence 21919, A
22	178	8.2	500	13	US-10-085-783A-20597	Sequence 20597, A
23	178	8.2	500	16	US-10-242-535A-20597	Sequence 20597, A
24	160	7.3	719	13	US-10-027-632-173199	Sequence 173199,
25	160	7.3	719	16	US-10-027-632-173199	Sequence 173199,
c 26	153	7.0	740	13	US-10-027-632-149888	Sequence 149888,
c 27	153	7.0	740	16	US-10-027-632-149888	Sequence 149888,
c 28	113	5.2	492	13	US-10-240-425-1387	Sequence 1387, Ap
29	109	5.0	707	17	US-10-641-643-209	Sequence 209, App
30	79	3.6	572	13	US-10-027-632-244176	Sequence 244176,
31	79	3.6	572	16	US-10-027-632-244176	Sequence 244176,
32	74	3.4	325	13	US-10-085-783A-54629	Sequence 54629, A
33	74	3.4	325	16	US-10-242-535A-54629	Sequence 54629, A
34	60	2.8	60	10	US-09-908-975-16348	Sequence 16348, A
35	40	1.8	752	9	US-09-765-231A-79	Sequence 79, Appl
c 36	25	1.1	25	12	US-10-332-929A-7	Sequence 7, Appli
37	23	1.1	23	12	US-10-332-929A-9	Sequence 9, Appli
c 38	22	1.0	22	12	US-10-332-929A-8	Sequence 8, Appli
39	22	1.0	22	12	US-10-332-929A-10	Sequence 10, Appl
c 40	22	1.0	440	13	US-10-424-599-46277	Sequence 46277, A
c 41	22	1.0	2656	13	US-10-425-114-15719	Sequence 15719, A
c 42	21	1.0	21	12	US-10-332-929A-4	Sequence 4, Appli
43	21	1.0	632	13	US-10-027-632-270863	Sequence 270863,
44	21	1.0	632	16	US-10-027-632-270863	Sequence 270863,
c 45	21	1.0	733	13	US-10-027-632-265089	Sequence 265089,

ALIGNMENTS

RESULT 1
US-10-332-929A-17
; Sequence 17, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 2089
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-332-929A-17

Query Match 91.6%; Score 1994; DB 12; Length 2089;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy      193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy      253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy      313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284
```

```
Qy       373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
Db       285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy       433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
Db       345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy       493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
Db       405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy       553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
Db       465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy       613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
Db       525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy       673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
Db       585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy       733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
Db       645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy       793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
Db       705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy       853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
Db       765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy       913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
Db       825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy       973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
Db       885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy      1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
Db       945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy      1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
Db      1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy      1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
Db      1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy      1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
Db      1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy      1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
Db      1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy      1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
Db      1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy      1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
Db      1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy      1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
Db      1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy      1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
Db      1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy      1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
Db      1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy      1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
Db      1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy      1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
Db      1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy      1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
Db      1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy      1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
Db      1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy      1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
Db      1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy      1933 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1992
Db      1845 CCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCC 1904

Qy      1993 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 2052
Db      1905 TTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAG 1964

Qy      2053 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2112
Db      1965 GAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAA 2024

Qy      2113 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2172
Db      2025 TACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAA 2084

Qy      2173 AAAAA 2177
Db      2085 AAAAA 2089
```

```
RESULT 2
US-10-332-929A-1
; Sequence 1, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 2589
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2498)..(2498)
;   OTHER INFORMATION: n is a, c, g, or t
```

US-10-332-929A-1

Query Match 80.8%; Score 1759; DB 12; Length 2589;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1809; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
Db        45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy       193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
Db       105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy       253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
Db       165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy       313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
Db       225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy       373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
Db       285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy       433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
Db       345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy       493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
Db       405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy       553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
Db       465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy       613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
Db       525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy       673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
Db       585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy       733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
Db       645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy       793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
Db       705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy       853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
Db       765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy       913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
Db       825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy       973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
Db       885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy      1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
Db       945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy      1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy      1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy      1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy      1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy      1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy      1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy      1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy      1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy      1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy      1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy      1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy      1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy      1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy      1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy      1933 CCTGCGTGGG 1942
Db       1845 CCTGCGTGGG 1854
```

RESULT 3
US-10-332-929A-16
; Sequence 16, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 2039
; TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-10-332-929A-16

  Query Match             80.8%;  Score 1758;  DB 12;  Length 2039;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
              |||||||||
Db       1845 CCTGCGTGG 1853
```

```
RESULT 4
US-10-332-929A-18
; Sequence 18, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 2145
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-332-929A-18

  Query Match             80.8%;  Score 1758;  DB 12;  Length 2145;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1808;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        133 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 GCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 104

Qy        193 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        105 TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGC 164

Qy        253 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 CTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGG 224

Qy        313 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 ACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGA 284

Qy        373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 344

Qy        433 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACG 492
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        345 ACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACG 404

Qy        493 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 GGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCA 464

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        465 CGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATG 524

Qy        613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        525 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGG 584

Qy        673 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        585 ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGG 644

Qy        733 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        645 GCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGG 704

Qy        793 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        705 CGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA 764

Qy        853 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        765 GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGC 824

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        825 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 884

Qy        973 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        885 GTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATG 944

Qy       1033 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        945 GCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGA 1004

Qy       1093 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1005 GCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCC 1064

Qy       1153 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1065 CTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCA 1124

Qy       1213 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1125 ACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGC 1184

Qy       1273 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1185 ACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGG 1244

Qy       1333 GCACAGGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1245 GCACAGGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCC 1304

Qy       1393 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1305 ATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACA 1364

Qy       1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1365 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1424

Qy       1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1425 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1484

Qy       1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1485 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1544

Qy       1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1545 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1604

Qy       1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1605 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1664

Qy       1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1665 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1724

Qy       1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1725 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCT 1784

Qy       1873 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1785 ACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAG 1844

Qy       1933 CCTGCGTGG 1941
              |||||||||
Db       1845 CCTGCGTGG 1853


RESULT 5
US-10-295-027-182
; Sequence 182, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
```

; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 182
; LENGTH: 2507
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-182

Query Match 80.2%; Score 1745; DB 16; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||| .
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796


RESULT 6
US-10-188-832-57
; Sequence 57, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
;  APPLICANT: Mack, David H.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
;  TITLE OF INVENTION:  and Methods of Screening for Modulators of Bladder
;  TITLE OF INVENTION:  Cancer
;  FILE REFERENCE: 018501-002330US
;  CURRENT APPLICATION NUMBER: US/10/188,832
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US 60/302,814
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: US 60/310,099
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/343,705
;  PRIOR FILING DATE: 2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/372,246
;  PRIOR FILING DATE: 2002-04-12
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
;   LENGTH: 2507
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-188-832-57

  Query Match             80.2%;  Score 1745;  DB 17;  Length 2507;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||

Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
```

```
Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 7
US-09-822-830A-88
; Sequence 88, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 88
; LENGTH: 1849
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1810,1813,1814
; OTHER INFORMATION: n=a,c,g, or t
US-09-822-830A-88

Query Match 71.9%; Score 1565; DB 9; Length 1849;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1665; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 60

Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 120

Qy        359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 180

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGC 240

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 300

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 360

Qy        599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 420

Qy        659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 480

Qy        719 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 540

Qy        779 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 600

Qy        839 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 660

Qy        899 CAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 958
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGAGAATGGGCCTAAATTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 720

Qy        959 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 780

Qy       1019 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 840

Qy       1079 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 900

Qy       1139 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 960

Qy       1199 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1020

Qy       1259 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1080

Qy       1319 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1140

Qy       1379 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1200

Qy       1439 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1260
```

```
Qy       1499 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1320

Qy       1559 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1380

Qy       1619 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1440

Qy       1679 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1500

Qy       1739 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1798
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1560

Qy       1799 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1858
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1620

Qy       1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGG 1905
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGG 1667


RESULT 8
US-09-765-231A-30
; Sequence 30, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
;  APPLICANT: Searle/Monsanto
;  APPLICANT:  Phippard, Deborah
;  APPLICANT:  Vasanthakamur, Geetha
;  APPLICANT:  Dotson, Stanton
;  APPLICANT:  Ma, Xiao-Jun
;  TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,
;  TITLE OF INVENTION:  vectors, and cells
;  FILE REFERENCE: SO-3221 PR
;  CURRENT APPLICATION NUMBER: US/09/765,231A
;  CURRENT FILING DATE:  2001-01-18
;  NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 30
;   LENGTH: 2263
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: 1-2263
;   LOCATION: unknown
;   OTHER INFORMATION: unsure at all n locations
US-09-765-231A-30

  Query Match             65.6%;  Score 1428;  DB 9;  Length 2263;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 212

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 272

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 332

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 392

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 452

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 512

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 572

Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 632

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 692

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 752

Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 812

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 872

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 932

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 992

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1052

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1112

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1172

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1232

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1292

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1352

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1412

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1472
```

```
Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1532

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1592

Qy       1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1517
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652

Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712

Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772

Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832

Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1757
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1892

Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952

Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012

Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072

Qy       1937 CGTGG 1941
             |||||
Db       2073 CGTGG 2077
```

```
RESULT 9
US-10-037-270-1032
; Sequence 1032, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1032
;   LENGTH: 1501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (161)..(1468)
US-10-037-270-1032

  Query Match             60.4%;  Score 1314;  DB 15;  Length 1501;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1314;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         36 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 95

Qy        688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 155

Qy        748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 215

Qy        808 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 275

Qy        868 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 335

Qy        928 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 395

Qy        988 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 455

Qy       1048 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 515

Qy       1108 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 575

Qy       1168 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 635

Qy       1228 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 695

Qy       1288 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 755

Qy       1348 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 815

Qy       1408 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        816 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 875

Qy       1468 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 1527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 935

Qy       1528 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995

Qy       1588 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        996 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055

Qy       1648 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1707
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1115

Qy       1708 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1767
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1116 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1175

Qy       1768 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1176 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1235

Qy       1828 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1236 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1295

Qy       1888 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1296 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1349


RESULT 10
US-10-117-722-1032
; Sequence 1032, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2BCIP
;  CURRENT APPLICATION NUMBER: US/10/117,722
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: 09/620,312
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1032
;   LENGTH: 1501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (161)..(1468)
US-10-117-722-1032

  Query Match             60.4%;  Score 1314;  DB 16;  Length 1501;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1314;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        628 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 687
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         36 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 95

Qy        688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 155

Qy        748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 215

Qy        808 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 275

Qy        868 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCC 335

Qy        928 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 ACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCC 395

Qy        988 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 1047
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCG 455

Qy       1048 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 1107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 TCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGG 515

Qy       1108 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 1167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 TCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCA 575

Qy       1168 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 1227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 TCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACC 635

Qy       1228 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 1287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCA 695

Qy       1288 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 1347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        696 TCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGG 755

Qy       1348 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 1407
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 TGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGG 815

Qy       1408 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 1467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        816 CTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 875

Qy       1468 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 1527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        876 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 935

Qy       1528 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        936 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 995

Qy       1588 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        996 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAG 1055

Qy       1648 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1056 ATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCC 1115

Qy       1708 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1116 TCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGAT 1175
```

```
Qy       1768 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1176 TCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCG 1235

Qy       1828 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1236 TGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1295

Qy       1888 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1296 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1349


RESULT 11
US-10-295-027-180
; Sequence 180, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 180
;   LENGTH: 2178
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-295-027-180

  Query Match             56.8%;  Score 1236;  DB 16;  Length 2178;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1286;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        468 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 93

Qy        528 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 587
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         94 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 153

Qy        588 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 647
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 213

Qy        648 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 707
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 273

Qy        708 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 767
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 333

Qy        768 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 827
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 393

Qy        828 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 887
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 453

Qy        888 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 947
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 513

Qy        948 TTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTC 1007
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 TTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTC 573

Qy       1008 GCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGC 1067
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGC 633

Qy       1068 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 1127
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 693

Qy       1128 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAAT 1187
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAAT 753

Qy       1188 GACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGC 1247
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 GACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGC 813

Qy       1248 CTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGC 1307
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 CTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGC 873

Qy       1308 GACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGAC 1367
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 GACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGAC 933

Qy       1368 GGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTC 1427
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTC 993

Qy       1428 CGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGG 1487
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        994 CGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGGTTGGG 1053

Qy       1488 GCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGG 1547
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 GCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGG 1113

Qy       1548 ATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTG 1607
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 ATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTG 1173

Qy       1608 GGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGG 1667
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 GGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGG 1233
```

```
Qy       1668 AACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGAC 1727
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 AACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGAC 1293

Qy       1728 ACACTTCAGGACCCAGCCCCACTGGAG 1754
             |||||||||||||||||||||||||||
Db       1294 ACACTTCAGGACCCAGCCCCACTGGAG 1320


RESULT 12
US-10-188-832-55
; Sequence 55, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
;  APPLICANT: Mack, David H.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
;  TITLE OF INVENTION:  and Methods of Screening for Modulators of Bladder
;  TITLE OF INVENTION:  Cancer
;  FILE REFERENCE: 018501-002330US
;  CURRENT APPLICATION NUMBER: US/10/188,832
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US 60/302,814
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: US 60/310,099
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/343,705
;  PRIOR FILING DATE: 2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/372,246
;  PRIOR FILING DATE: 2002-04-12
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
;   LENGTH: 2178
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-188-832-55

  Query Match             56.8%;  Score 1236;  DB 17;  Length 2178;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1286;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        468 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 93

Qy        528 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 153

Qy        588 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 213

Qy        648 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 273

Qy        708 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 767
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 333

Qy        768 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 393

Qy        828 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 453

Qy        888 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 513

Qy        948 TTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTC 1007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 TTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTC 573

Qy       1008 GCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGC 1067
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGC 633

Qy       1068 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 1127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 693

Qy       1128 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAAT 1187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAAT 753

Qy       1188 GACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGC 1247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 GACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGC 813

Qy       1248 CTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGC 1307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 CTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGC 873

Qy       1308 GACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGAC 1367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 GACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGAC 933

Qy       1368 GGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTC 1427
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTC 993

Qy       1428 CGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGG 1487
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        994 CGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGGTTGGG 1053

Qy       1488 GCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGG 1547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1054 GCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGG 1113

Qy       1548 ATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTG 1607
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1114 ATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTG 1173

Qy       1608 GGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGG 1667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1174 GGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGG 1233

Qy       1668 AACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGAC 1727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1234 AACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGAC 1293

Qy       1728 ACACTTCAGGACCCAGCCCCACTGGAG 1754
              |||||||||||||||||||||||||||
Db        1294 ACACTTCAGGACCCAGCCCCACTGGAG 1320


RESULT 13
US-10-641-643-484
; Sequence 484, Application US/10641643
; Publication No. US20040077003A1
;  GENERAL INFORMATION:
;       APPLICANT: Cocks, Benjamin G.
;                  Susan G. Stuart
;                  Jeffrey J. Seilhamer
;       TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
;                           GENE EXPRESSION
;       NUMBER OF SEQUENCES: 1508
```

```
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;            STREET: 3174 PORTER DRIVE
;            CITY: PALO ALTO
;            STATE: CALIFORNIA
;            COUNTRY: USA
;            ZIP: 94304
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/641,643
;            FILING DATE: 14-Aug-2003
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: <Unknown>
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Zeller, Karen J.
;            REGISTRATION NUMBER: 37,071
;            REFERENCE/DOCKET NUMBER: PA-0001 US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (650) 855-0555
;            TELEFAX: (650) 845-4166
;   INFORMATION FOR SEQ ID NO: 484:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 789 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       IMMEDIATE SOURCE:
;            LIBRARY: SYNORAB01
;            CLONE: 192279
;       SEQUENCE DESCRIPTION: SEQ ID NO: 484 :
US-10-641-643-484

  Query Match             33.9%;  Score 738;  DB 17;  Length 789;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  788;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1058 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 1117
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 60

Qy       1118 GGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 1177
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGACATCGNCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 120

Qy       1178 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 180

Qy       1238 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 240

Qy       1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 300

Qy       1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 360

Qy       1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 420

Qy       1478 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 480

Qy       1538 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 540

Qy       1598 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 1657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 600

Qy       1658 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 660

Qy       1718 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 1777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 720

Qy       1778 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 1837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 780

Qy       1838 AGACAAGCC 1846
             |||||||||
Db         781 AGACAAGCC 789


RESULT 14
US-10-295-027-184
; Sequence 184, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
;   LENGTH: 4794
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

US-10-295-027-184

Query Match 31.6%; Score 687; DB 16; Length 4794;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1256 CGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTT 1315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3891 CGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTT 3950

Qy       1316 GGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCT 1375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3951 GGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCT 4010

Qy       1376 GGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAA 1435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4011 GGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAA 4070

Qy       1436 TCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGC 1495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4071 TCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGC 4130

Qy       1496 CAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGA 1555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4131 CAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGA 4190

Qy       1556 CGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGA 1615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4191 CGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGA 4250

Qy       1616 TGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGC 1675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4251 TGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGC 4310

Qy       1676 CAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCA 1735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4311 CAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCA 4370

Qy       1736 GGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCAT 1795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4371 GGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCAT 4430

Qy       1796 GGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGT 1855
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4431 GGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGT 4490

Qy       1856 CAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCC 1915
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4491 CAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCC 4550

Qy       1916 CAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||
Db       4551 CAACGAGGATGGCACAGCCTGCGTGGG 4577
```

RESULT 15
US-10-188-832-59
; Sequence 59, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 4794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-188-832-59

Query Match 31.6%; Score 687; DB 17; Length 4794;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1256 CGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTT 1315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3891 CGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTT 3950

Qy       1316 GGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCT 1375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3951 GGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCT 4010

Qy       1376 GGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAA 1435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4011 GGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAA 4070

Qy       1436 TCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGC 1495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4071 TCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGC 4130

Qy       1496 CAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGA 1555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4131 CAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGA 4190

Qy       1556 CGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGA 1615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4191 CGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGA 4250

Qy       1616 TGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGC 1675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4251 TGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGC 4310

Qy       1676 CAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCA 1735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4311 CAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCA 4370

Qy       1736 GGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCAT 1795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4371 GGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCAT 4430

Qy       1796 GGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGT 1855
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4431 GGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGT 4490

Qy       1856 CAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCC 1915
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4491 CAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCC 4550

Qy       1916 CAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||
Db       4551 CAACGAGGATGGCACAGCCTGCGTGGG 4577
```

Search completed: August 10, 2004, 08:07:17
Job time : 5602 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 23:06:56 ; Search time 5087 Seconds
(without alignments)
12779.628 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1438	66.1	2147	11	BC042687	BC042687 Mus muscu
2	719	33.0	781	12	BI550776	BI550776 603195670
3	653	30.0	841	12	BG708845	BG708845 602673155
c 4	632	29.0	656	13	BQ183029	BQ183029 UI-H-EU0-
5	626	28.8	636	10	BF727258	BF727258 by19a09.y
6	613	28.2	1101	13	BQ067249	BQ067249 AGENCOURT
7	590	27.1	770	12	BI758703	BI758703 603024078
8	566	26.0	926	10	BF527664	BF527664 602040644
9	553	25.4	904	13	BX451456	BX451456 BX451456
10	519	23.8	735	12	BG911024	BG911024 602813023
11	516	23.7	567	12	BM707643	BM707643 UI-E-CI1-
12	514	23.6	1063	12	BI596960	BI596960 603242976
13	507	23.3	578	14	CD676152	CD676152 fs32g06.y
14	501	23.0	804	12	BI458314	BI458314 603199270
15	496	22.8	608	9	AU123536	AU123536 AU123536
16	488	22.4	540	14	CB153273	CB153273 K-EST0210
17	486	22.3	598	13	BQ637467	BQ637467 he10e02.y
18	483	22.2	566	12	BG721807	BG721807 602695986
19	477	21.9	511	9	AI791928	AI791928 ag11c02.y
20	473	21.7	551	12	BM699706	BM699706 UI-E-DW1-
21	466	21.4	489	10	BF725127	BF725127 bx12e04.y
22	458	21.0	580	12	BI549399	BI549399 603190054
23	455	20.9	1022	13	BX460454	BX460454 BX460454
24	451	20.7	713	9	AI792073	AI792073 ag62d02.y
25	447	20.5	539	14	CB152752	CB152752 K-EST0210
26	442	20.3	477	13	BX102537	BX102537 BX102537
27	411	18.9	622	12	BI824952	BI824952 603033366
28	409	18.8	622	12	BG716170	BG716170 602677669
29	399	18.3	399	14	CA388867	CA388867 cs03e01.y
30	394	18.1	549	12	BM706915	BM706915 UI-E-CQ0-
c 31	383	17.6	533	10	BE503986	BE503986 hz54c06.x
32	369	16.9	430	10	BF930592	BF930592 MR2-NT013
33	357	16.4	508	10	BF088585	BF088585 RC1-HT088
34	348	16.0	433	12	BG900984	BG900984 HOA52-1-D
35	337	15.5	860	12	BI819378	BI819378 603034549
36	327	15.0	438	12	BG899412	BG899412 HOA45-1-A
37	326	15.0	338	10	BF727076	BF727076 by16d07.y
38	313	14.4	398	12	BG898358	BG898358 HOA55-1-E
c 39	305	14.0	470	10	BF727075	BF727075 by16d07.x
40	305	14.0	729	12	BG818593	BG818593 602779443
c 41	295	13.6	350	10	BF736863	BF736863 QV3-KT001
c 42	272	12.5	470	10	BE503764	BE503764 hz51f03.x
43	237	10.9	539	10	BF849175	BF849175 QV0-EN005
44	226	10.4	418	14	T77333	T77333 yc93h03.r1
45	218	10.0	380	12	BG820409	BG820409 602780470

ALIGNMENTS

RESULT 1
BC042687
LOCUS BC042687 2147 bp mRNA linear HTC 02-JAN-2003
DEFINITION Mus musculus, clone IMAGE:3157049, mRNA.
ACCESSION BC042687
VERSION BC042687.1 GI:27469363
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2147)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC

Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 6 Row: p Column: 3
This clone has the following problem: no 5' EST match.

FEATURES Location/Qualifiers
source 1. .2147
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3157049"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 66.1%; Score 1438; DB 11; Length 2147;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
Db         1 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 60
Qy       565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
Db        61 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 120
Qy       625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
Db       121 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 180
Qy       685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
Db       181 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 240
Qy       745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
Db       241 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 300
Qy       805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
Db       301 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 360
Qy       865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
Db       361 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 420
Qy       925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
Db       421 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 480
Qy       985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
Db       481 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 540
Qy      1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
Db       541 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 600
Qy      1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
Db       601 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 660
Qy      1165 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 1224
Db       661 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 720
Qy      1225 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 1284
Db       721 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 780
Qy      1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
Db       781 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 840
Qy      1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
Db       841 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 900
Qy      1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
Db       901 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 960
Qy      1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
Db       961 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1020
Qy      1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
Db      1021 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1080
Qy      1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
Db      1081 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1140
Qy      1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
Db      1141 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1200
Qy      1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
Db      1201 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1260
Qy      1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
Db      1261 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1320
Qy      1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
Db      1321 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1380
Qy      1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
Db      1381 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1438
```

RESULT 2
BI550776
LOCUS BI550776 781 bp mRNA linear EST 05-SEP-2001
DEFINITION 603195670F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5275168 5', mRNA sequence.
ACCESSION BI550776
VERSION BI550776.1 GI:15438088
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 781)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11694  row: 1  column: 17
           High quality sequence stop: 763.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5275168"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             33.0%;  Score 719;  DB 12;  Length 781;
  Best Local Similarity   99.9%;  Pred. No. 5e-240;
  Matches  769;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 603

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 663

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 723

Qy        723 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 773


RESULT 3
BG708845
LOCUS       BG708845                841 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602673155F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4795957 5',
            mRNA sequence.
ACCESSION   BG708845
VERSION     BG708845.1  GI:13986590
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10679  row: m  column: 14
           High quality sequence stop: 822.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4795957"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length  clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             30.0%;  Score 653;  DB 12;  Length 841;
  Best Local Similarity   99.9%;  Pred. No. 4.6e-217;
  Matches  703;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123
```

```
Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 603

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 663

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACAT 706
              |||||||||||||||||||||||||||||||||||||||||||
Db        664 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACAT 707
```

RESULT 4
BQ183029/c

```
LOCUS       BQ183029                656 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  UI-H-EU0-azs-e-03-0-UI.s1 NCI_CGAP_Car1 Homo sapiens cDNA clone
            IMAGE:5852906 3', mRNA sequence.
ACCESSION   BQ183029
VERSION     BQ183029.1  GI:20358579
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5852906"
                     /tissue_type="Osteoarthritic Cartilage"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Car1"
                     /note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_Car1 is a cDNA library containing the following
                     tissue(s): Osteoarthritic Cartilage The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_TISSUE=osteoarthritic cartilage
                     TAG_LIB=UI-H-EU0
                     TAG_SEQ=TGATCACGCT"
ORIGIN
```

```
  Query Match           29.0%;  Score 632;  DB 13;  Length 656;
  Best Local Similarity  100.0%;  Pred. No. 1e-209;
  Matches  632;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1534 GGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGG 1593
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGG 592

Qy       1594 CACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCA 1653
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         591 CACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCA 532

Qy       1654 AGATGGTGAGCCCGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACC 1713
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         531 AGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACC 472

Qy       1714 CCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCC 1773
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         471 CCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCC 412

Qy       1774 AGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCC 1833
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         411 AGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCC 352

Qy       1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 1893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         351 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 292

Qy       1894 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 1953
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         291 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 232

Qy       1954 CTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAAT 2013
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         231 CTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAAT 172

Qy       2014 CACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCA 2073
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         171 CACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCA 112

Qy       2074 CTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAATACATTAAGATTTGGGGGTGC 2133
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         111 CTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAATACATTAAGATTTGGGGGTGC 52

Qy       2134 TACCTTACATAATAAATTCCCATTTCCTCTTG 2165
             ||||||||||||||||||||||||||||||||
Db          51 TACCTTACATAATAAATTCCCATTTCCTCTTG 20
```

```
RESULT 5
BF727258
LOCUS       BF727258                636 bp    mRNA    linear   EST 05-JAN-2001
DEFINITION  by19a09.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by19a09 5', mRNA sequence.
ACCESSION   BF727258
VERSION     BF727258.1  GI:12043169
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 19  row: a  column: 09
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="by19a09"
                     /tissue_type="Lens"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                     BY"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                     from different adults (both approximately 40 years old)
                     together yielded 20ug of total RNA and 150ng mRNA for cDNA
                     library synthesis. A directionally cloned cDNA library in
                     the pCMVSPORT6 vector was constructed at Life
                     Technologies, essentially following the protocols of the
                     SuperScript Plasmid System full details of which are
                     contained in the manufacturer's Instruction manual
                     (http://www.lifetech.com/). First strand synthesis was
                     carried out using a Not I primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                     end inserts were cloned into the Not I/EcoR V sites in the
                     vector. EST analysis was performed on the unamplified
                     library at the NIH Intramural Sequencing Center (NISC)."
ORIGIN

  Query Match             28.8%;  Score 626;  DB 10;  Length 636;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-207;
  Matches  626;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1277 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 1336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 70

Qy       1337 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGG 1396
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGG 130

Qy       1397 AGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTG 1456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 AGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTG 190

Qy       1457 GCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCT 1516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCT 250

Qy       1517 CTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTG 1576
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 CTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTG 310

Qy       1577 TGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGT 1636
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 TGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGT 370

Qy       1637 GACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGT 1696
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 GACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGT 430

Qy       1697 GCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTG 1756
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        431 GCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTG 490

Qy       1757 TGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCA 1816
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        491 TGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCA 550

Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        551 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 610

Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTC 1902
             |||||||||||||||||||||||||||
Db        611 GTGCCGGACCAACAAGAAGTGCAGTC 636


RESULT 6
BQ067249
LOCUS       BQ067249                1101 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6767325 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5751633
            5', mRNA sequence.
ACCESSION   BQ067249
VERSION     BQ067249.1  GI:19896295
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12784  row: i  column: 10
            High quality sequence stop: 612.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5751633"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_115"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
```

```
ORIGIN

  Query Match             28.2%;  Score 613;  DB 13;  Length 1101;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-203;
  Matches  613;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        720 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 60

Qy        780 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 120

Qy        840 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 180

Qy        900 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 240

Qy        960 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 300

Qy       1020 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 360

Qy       1080 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 420

Qy       1140 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 480

Qy       1200 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 540

Qy       1260 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 600

Qy       1320 CCTGAGGGCCGGG 1332
              |||||||||||||
Db        601 CCTGAGGGCCGGG 613


RESULT 7
BI758703
LOCUS       BI758703                 770 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603024078F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194530 5',
            mRNA sequence.
ACCESSION   BI758703
VERSION     BI758703.1  GI:15750294
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11486  row: 1  column: 19
            High quality sequence start: 3
            High quality sequence stop: 770.
FEATURES             Location/Qualifiers
     source          1. .770
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194530"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
ORIGIN

  Query Match             27.1%;  Score 590;  DB 12;  Length 770;
  Best Local Similarity   99.8%;  Pred. No. 4.1e-195;
  Matches  640;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 168

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 228

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 288

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 348

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 408

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        409 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCT 468

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 528

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 588

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 648

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 708

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCT 761
              |||||||||||||||||||||||||||||||||||||||||
Db        709 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCT 749
```

```
RESULT 8
BF527664
LOCUS       BF527664                926 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  602040644F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4178363
            5', mRNA sequence.
ACCESSION   BF527664
VERSION     BF527664.1  GI:11615027
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 926)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9486  row: h  column: 12
            High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1. .926
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4178363"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn67"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             26.0%;  Score 566;  DB 10;  Length 926;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-187;
  Matches  566;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 60

Qy        565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 120

Qy        625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 180

Qy        685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 240

Qy        745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 300

Qy        805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 360

Qy        865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 420

Qy        925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 480

Qy        985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 540

Qy       1045 TCGTCTATGGCAACTGGAATGGCCCC 1070
              |||||||||||||||||||||||||||
Db        541 TCGTCTATGGCAACTGGAATGGCCCC 566


RESULT 9
BX451456
LOCUS       BX451456                904 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX451456 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF014YH11 5-PRIME, mRNA sequence.
ACCESSION   BX451456
VERSION     BX451456.1  GI:31022247
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 904)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5247.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAF012ZE12_AF01150_1&cluster=5247.r.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAF012ZE12_AF01150_1.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF014YH11"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             25.4%;  Score 553;  DB 13;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-182;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        929 CAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCA 988
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 CAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCA 149

Qy        989 CCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGT 1048
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 CCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGT 209

Qy       1049 CTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGT 1108
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          210 CTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGAAGGT 269

Qy         1109 CCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCAT 1168
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          270 CCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCAT 329

Qy         1169 CACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCG 1228
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          330 CACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCG 389

Qy         1229 CAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCAT 1288
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          390 CAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCAT 449

Qy         1289 CGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGT 1348
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          450 CGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGT 509

Qy         1349 GACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGC 1408
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          510 GACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGC 569

Qy         1409 TCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGT 1468
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          570 TCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGT 629

Qy         1469 GCCACGCACCCGG 1481
                |||||||||||||
Db          630 GCCACGCACCCGG 642
```

```
RESULT 10
BG911024
LOCUS       BG911024                 735 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  602813023F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944964
            5', mRNA sequence.
ACCESSION   BG911024
VERSION     BG911024.1  GI:14291500
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10890  row: n  column: 05
            High quality sequence stop: 656.
FEATURES             Location/Qualifiers
     source          1..735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4944964"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn67"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             23.8%;  Score 519;  DB 12;  Length 735;
  Best Local Similarity   99.8%;  Pred. No. 2.2e-170;
  Matches  569;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 60

Qy          259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 120

Qy          319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 180

Qy          379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 240

Qy          439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
                |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          241 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 300

Qy          499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 360

Qy          559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 420

Qy          619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 480

Qy          679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 540

Qy          739 ATGCCCTCATTGAAATGGACCCTGAGGCCA 768
                ||||||||||||||||||||||||||||||
Db          541 ATGCCCTCATTGAAATGGACCCTGAGGCCA 570
```

```
RESULT 11
BM707643
LOCUS       BM707643                 567 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CI1-afs-o-03-0-UI.r1 UI-E-CI1 Homo sapiens cDNA clone
            UI-E-CI1-afs-o-03-0-UI 5', mRNA sequence.
ACCESSION   BM707643
VERSION     BM707643.1  GI:19020901
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
```

```
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CI1-afs-o-03-0-UI"
                     /tissue_type="RPE and Choroid"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-CI1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CI1 is a normalized cDNA library containing the
                     following tissue(s): RPE and Choroid. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is ACCTA.
                     This library was created for the program, Gene Discovery
                     in the Visual System, supported by National Eye Institute
                     (NEI)."
ORIGIN

  Query Match             23.7%;  Score 516;  DB 12;  Length 567;
  Best Local Similarity   99.8%;  Pred. No. 2.7e-169;
  Matches  566;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        880 CCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCG 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCG 60

Qy        940 ATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGC 999
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGC 120

Qy       1000 GAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACT 1059
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACT 180

Qy       1060 GGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGG 1119
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 GGAATGGCCCCCACCGCCTCTATCTGCANATGAGCACCCATGGGAAGGTCCGCTTCCGGG 240

Qy       1120 ACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACT 1179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACT 300

Qy       1180 TTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAG 1239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAG 360

Qy       1240 CCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCA 1299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCA 420

Qy       1300 ATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCG 1359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCG 480

Qy       1360 ACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGT 1419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 ACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGT 540

Qy       1420 CCGTCTTCCGGGGCAATCAGGGCTTCA 1446
              |||||||||||||||||||||||||||
Db        541 CCGTCTTCCGGGGCAATCAGGGCTTCA 567


RESULT 12
BI596960
LOCUS       BI596960                1063 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603242976F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5285419 5',
            mRNA sequence.
ACCESSION   BI596960
VERSION     BI596960.1  GI:15489899
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1063)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11721  row: g  column: 20
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .1063
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5285419"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             23.6%;  Score 514;  DB 12;  Length 1063;
  Best Local Similarity   99.8%;  Pred. No. 1e-168;
  Matches  564;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGT 567
                |||||||||||||||||||||||||
Db        544 GGGGTGGCCACGTACACCGACAAGT 568


RESULT 13
CD676152
LOCUS       CD676152                578 bp    mRNA    linear   EST 24-JUN-2003
DEFINITION  fs32g06.y1 Human Lens cDNA (Normalized): fs Homo sapiens cDNA clone
            fs32g06 5', mRNA sequence.
ACCESSION   CD676152
VERSION     CD676152.1  GI:32177883
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Wistow,G., Bernstein,S.L., Wyatt,M.K., Behal,A., Touchman,J.W.,
            Bouffard,G., Smith,D. and Peterson,K.
  TITLE     Expressed sequence tag analysis of adult human lens for the NEIBank
            Project: over 2000 non-redundant transcripts, novel genes and
            splice variants
  JOURNAL   Mol. Vis. 8 (4), 171-184 (2002)
  MEDLINE   22103463
   PUBMED   12107413
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 32  row: g  column: 06
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="fs32g06"
                     /tissue_type="Lens"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Lens cDNA (Normalized): fs"
                     /note="Organ: Eye; Vector: pCMVSPORT6; A human lens
                     library (by) was normalized by self-subtraction. One
                     portion of double stranded plasmid DNA representing the
                     library was linearized by NotI. This NotI digested library
                     was used as a template for biotinylated RNA synthesis
                     using SP6 RNA polymerase. Another portion of the double
                     stranded plasmid library was converted to single-stranded
                     circles in vitro using Gene II and Exonuclease III (Life
                     Technologies). Single-stranded DNA (1 mg) was hybridized
                     (C0t 500) with 41 mg of Bio-RNA and vector blocking
                     oligonucleotides. The hybridized Bio-RNA/ss-circles were
                     removed by streptavidin:phenol extraction. EST analysis
                     was performed on the library at the NIH Intramural
                     Sequencing Center(NISC)."
ORIGIN

  Query Match           23.3%;  Score 507;  DB 14;  Length 578;
  Best Local Similarity   99.8%;  Pred. No. 3.6e-166;
  Matches  557;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         12 AGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGC 71
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGC 60

Qy         72 AGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCG 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCG 120

Qy        132 AGCGCTGACCCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCC 191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGCGCTGACCCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCC 180

Qy        192 ATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTG 251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTG 240

Qy        252 CCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTG 311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTG 300

Qy        312 GACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTG 371
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTG 360

Qy        372 AAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCC 431
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCC 420

Qy        432 TACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGAC 491
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        421 TACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGAC 480

Qy        492 GGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCC 551
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCC 540

Qy        552 ACGTACACCGACAAGTTG 569
              ||||||||||||||||||
Db        541 ACGTACACCGACAAGTTG 558


RESULT 14
BI458314
LOCUS       BI458314                804 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603199270F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5278624 5',
            mRNA sequence.
ACCESSION   BI458314
VERSION     BI458314.1  GI:15248970
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11703  row: 1  column: 17
            High quality sequence stop: 785.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5278624"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             23.0%;  Score 501;  DB 12;  Length 804;
  Best Local Similarity   99.5%;  Pred. No. 3.8e-164;
  Matches  651;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCA 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCA 61

Qy         64 GAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGA 121

Qy        124 TGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGT 181

Qy        184 TTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACT 241

Qy        244 CAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTA 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTA 301

Qy        304 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACC 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACC 361

Qy        364 TGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA 423
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        362 TGGTTCTGAAGTATGACCAGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA 421

Qy        424 GCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCG 483
              |||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        422 GCTCACACTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCG 481

Qy        484 ACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGG 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 ACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGG 541

Qy        544 GGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 GGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCC 601

Qy        604 TGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCT 657
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCT 655


RESULT 15
AU123536
LOCUS       AU123536                608 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU123536 NT2RM2 Homo sapiens cDNA clone NT2RM2000502 5', mRNA
            sequence.
ACCESSION   AU123536
VERSION     AU123536.1  GI:10948252
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
            Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .608
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM2000502"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
ORIGIN

  Query Match             22.8%;  Score 496;  DB 9;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-162;
  Matches  496;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        468 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 GGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACC 93

Qy        528 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 587
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         94 AATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAAC 153

Qy        588 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 647
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 CGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTT 213

Qy        648 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 707
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 GCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATT 273

Qy        708 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 767
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          274 GCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCC 333

Qy          768 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 827
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          334 AGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGC 393

Qy          828 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 887
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          394 AAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGAT 453

Qy          888 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 947
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          454 ATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC 513

Qy          948 TTTGTGGACGCTGCGG 963
                ||||||||||||||||
Db          514 TTTGTGGACGCTGCGG 529
```

Search completed: August 10, 2004, 03:32:25
Job time : 5093 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:36:29 ; Search time 100 Seconds
(without alignments)
12302.118 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161651_29663/app_query.fasta_1.2375
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914958_@CGN_1_1_81_@runat_03082004_161651_29663 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3335	82.2	626	3	AAB08864	Aab08864 Amino aci
2	3292	81.2	618	5	AAE18681	Aae18681 Human SCI
3	3190.5	78.7	653	5	AAO18264	Aao18264 Human cho
4	3190.5	78.7	653	5	AAE18679	Aae18679 Human SCI
5	3190.5	78.7	653	6	ABR48171	Abr48171 Human bla
6	3190.5	78.7	653	6	ABU56636	Abu56636 Lung canc
7	3181.5	78.4	637	5	AAE18680	Aae18680 Human SCI
8	3179	78.4	612	5	AAE18682	Aae18682 Human SCI
9	3137	77.3	636	5	AAO18267	Aao18267 Human cho
10	2337	57.6	1597	6	ABR48172	Abr48172 Human bla
11	2337	57.6	1597	6	ABU56637	Abu56637 Lung canc
12	2124.5	52.4	435	4	AAM39983	Aam39983 Human pol
13	1967.5	48.5	418	4	AAB92693	Aab92693 Human pro
14	1967.5	48.5	418	6	ABR48170	Abr48170 Human bla
15	1967.5	48.5	418	6	ABU56635	Abu56635 Lung canc
16	1492	36.8	306	3	AAY76247	Aay76247 Fragment
17	1492	36.8	306	7	ADE11877	Ade11877 Human sec
18	1039	25.6	229	5	ABB09746	Abb09746 Amino aci
19	465	11.5	98	4	AAM41769	Aam41769 Human pol
20	297	7.3	1212	2	AAW87503	Aaw87503 Human N-m
21	297	7.3	19938	6	ABP76679	Abp76679 Streptomy
22	295.5	7.3	1061	2	AAW87504	Aaw87504 Human N-m
23	292	7.2	1081	3	AAB26240	Aab26240 Human N-m
24	292	7.2	1081	5	ABB56523	Abb56523 Human NMD
25	292	7.2	1081	6	ABU61440	Abu61440 Human N-m
26	292	7.2	1081	7	AAE39251	Aae39251 Human NMD
c 27	288.5	7.0	19938	6	ABP76679	Abp76679 Streptomy
c 28	286	6.9	19938	6	ABP76680	Abp76680 Streptomy
29	285.5	7.0	1232	3	AAB26239	Aab26239 Human N-m
30	285.5	7.0	1232	5	AAM47961	Aam47961 Human NMD
31	285.5	7.0	1232	6	ABU61439	Abu61439 Human N-m
32	285.5	7.0	1232	7	AAE39250	Aae39250 Human NMD
33	274.5	6.8	4315	5	ABP43908	Abp43908 MUC5B par
c 34	273.5	6.6	19938	6	ABP76678	Abp76678 Streptomy
c 35	273.5	6.6	19938	6	ABP76682	Abp76682 Streptomy
36	270.5	6.7	19938	6	ABB98398	Abb98398 Streptomy
37	269.5	6.6	1093	2	AAR41001	Aar41001 Human myo
c 38	268	6.5	19938	6	ABB98398	Abb98398 Streptomy
39	267	6.6	50	6	ADA57202	Ada57202 Human sec
40	267	6.6	50	6	ADA41079	Ada41079 Human sec
41	267	6.6	50	7	ADE11753	Ade11753 Human sec
42	267	6.6	51	3	AAY76131	Aay76131 Human sec
43	267	6.6	19938	6	ABP76682	Abp76682 Streptomy
44	255.5	6.3	19938	6	ABP76678	Abp76678 Streptomy
45	253.5	6.2	695	4	ABU53152	Abu53152 Human tes

ALIGNMENTS

RESULT 1
AAB08864
ID AAB08864 standard; protein; 626 AA.
XX
AC AAB08864;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of a human secretory protein.
XX
KW Human; secretory protein; HSECP; cancer; gastrointestinal disorder;
KW inflammation; cardiovascular disorder; neurological disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1. .27
FT /note= "signal sequence"
FT Modified-site 28
FT /note= "potential phosphorylation site"
FT Modified-site 61
FT /note= "potential phosphorylation site"
FT Modified-site 118
FT /note= "potential phosphorylation site"
FT Modified-site 144
FT /note= "potential phosphorylation site"
FT Modified-site 146
FT /note= "potential phosphorylation site"
FT Modified-site 188
FT /note= "potential phosphorylation site"
FT Modified-site 252
FT /note= "potential phosphorylation site"
FT Modified-site 275
FT /note= "potential phosphorylation site"

```
FT   Modified-site   336
FT                   /note= "potential phosphorylation site"
FT   Modified-site   410
FT                   /note= "potential phosphorylation site"
FT   Modified-site   423
FT                   /note= "potential phosphorylation site"
FT   Modified-site   467
FT                   /note= "potential phosphorylation site"
FT   Modified-site   501
FT                   /note= "potential phosphorylation site"
FT   Modified-site   506
FT                   /note= "potential phosphorylation site"
FT   Modified-site   524
FT                   /note= "potential phosphorylation site"
FT   Modified-site   536
FT                   /note= "potential phosphorylation site"
FT   Modified-site   550
FT                   /note= "potential phosphorylation site"
FT   Modified-site   583
FT                   /note= "potential phosphorylation site"
FT   Modified-site   588
FT                   /note= "potential phosphorylation site"
FT   Modified-site   593
FT                   /note= "potential phosphorylation site"
XX
PN   WO200052151-A2.
XX
PD   08-SEP-2000.
XX
PF   03-MAR-2000; 2000WO-US005621.
XX
PR   05-MAR-1999;   99US-0123117P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Baughn MR,  Yue H,  Au-Young J,  Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-579282/54.
DR   N-PSDB; AAA75118.
XX
PT   Twenty two human secretory proteins for diagnosing, treating and
PT   preventing cancer, inflammation, and gastrointestinal, cardiovascular and
PT   neurological disorders.
XX
PS   Claim 1; Page 88-89; 107pp; English.
XX
CC   The present sequence represents a human secretory protein, designated
CC   HSECP-1. The specification also describes HSECP-2 to HSECP-22. The
CC   proteins are useful for diagnosing, treating and preventing cancer,
CC   inflammation, and gastrointestinal, cardiovascular and neurological
CC   disorders. The proteins may also be used to identify agonists,
CC   antagonists, and inhibitors. The polynucleotides may be used for
CC   producing the protein recombinantly, and as a source of probes and
CC   primers for isolating and identifying related sequences
XX
SQ   Sequence 626 AA;

Alignment Scores:
Pred. No.:                4.9e-231      Length:          626
Score:                    3335.00       Matches:         626
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              82.22%        Indels:          0
DB:                       3             Gaps:            0

US-09-914-958B-35 (1-2177) x AAB08864 (1-626)

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaProSerAlaAspProGlyMetSerArgMetLeuProPheLeuLeuLeuLeuTrp 20

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         21 PheLeuProIleThrGluGlySerGlnArgAlaGluProMetPheThrAlaValThrAsn 40

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerValLeuProProAspTyrAspSerAsnProThrGlnLeuAsnTyrGlyValAlaVal 60

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ThrAspValAspHisAspGlyAspPheGluIleValValAlaGlyTyrAsnGlyProAsn 80

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuValLeuLysTyrAspArgAlaGlnLysArgLeuValAsnIleAlaValAspGluArg 100

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAlaCys 120

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAsnAlaPheSer 140

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAspIle 160

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuSerAspGluValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSerVal 180

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaCysValAspArgLysGlySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyr 200

Qy        723 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGG 782
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyAsnValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSerArg 220

Qy        783 GGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGC 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrGlyGly 240

Qy        843 CGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAT 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgGlyValSerValGlyProIleLeuSerSerSerAlaSerAspIlePheCysAspAsn 260

Qy        903 GAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCG 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspGlyThrPheValAspAlaAla 280

Qy        963 GCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTC 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPhe 300

Qy       1023 AACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTAT 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyr 320

Qy       1083 CTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCC 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LeuGlnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSer 340

Qy       1143 ATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAG 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 MetProSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGlu 360

Qy       1203 ATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATC 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IlePhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgValIle 380

Qy       1263 CGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCT 1322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ArgArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGluPro 400
```

```
Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GluGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAspLeu 420

Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGC 1442
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGly 440

Qy       1443 TTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGA 1502
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 PheAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArgGly 460

Qy       1503 GCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGC 1562
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGly 480

Qy       1563 TCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCC 1622
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAla 500

Qy       1623 AGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGG 1682
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 SerSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGly 520

Qy       1683 GAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCA 1742
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GluMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspPro 540

Qy       1743 GCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACC 1802
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AlaProLeuGluCysGlyGlnGlyPheSerGlnGlnGluAsnGlyHisCysMetAspThr 560

Qy       1803 AATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACC 1862
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsnThr 580

Qy       1863 TATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAG 1922
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlySerTyrArgCysArgThrAsnLysLysCysSerArgGlyTyrGluProAsnGlu 600

Qy       1923 GATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTT 1982
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AspGlyThrAlaCysValGlyTrpTrpSerProValLeuLysIleValThrProGlnVal 620

Qy       1983 GGGAAGAGCCTTGGTCCC 2000
             ||||||||||||||||||
Db        621 GlyLysSerLeuGlyPro 626
```

RESULT 2
AAE18681
ID AAE18681 standard; protein; 618 AA.
XX
AC AAE18681;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 splice variant B protein.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy.
XX
OS Homo sapiens.
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR N-PSDB; AAD29792.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 6; Page 40-43; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for
CC induction of immunological tolerance to an autoantigen in patients
CC suffering from autoimmune disorders, more specifically rheumatoid
CC arthritis. The inflammatory disorders treated include diseases like
CC Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC multiple sclerosis and diabetes. Polypeptides of the invention are also
CC useful in a diagnostic method for the detection of activated autoreactive
CC T cells. The present sequence is human SCIM-1 splice variant B protein
XX
SQ Sequence 618 AA;

Alignment Scores:

Pred. No.:	6.03e-228	Length:	618
Score:	3292.00	Matches:	618
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	81.16%	Indels:	0
DB:	5	Gaps:	0

US-09-914-958B-35 (1-2177) x AAE18681 (1-618)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140
```

```
Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
Db       581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyTrp 600

Qy      1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
Db       601 TrpSerProValLeuLysIleValThrProGlnValGlyLysSerLeuGlyPro 618
```

RESULT 3
AAO18264
ID AAO18264 standard; protein; 653 AA.
XX
AC AAO18264;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein SEQ ID NO:4.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Peptide   1. .19
FT   /label= signal_peptide
FT Protein   20. .653
FT   /label= mature_CEP-68
XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR N-PSDB; AAL47917, AAL47918.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 1; Page 92-94; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is the CEP-68 protein
XX
SQ Sequence 653 AA;

```
Alignment Scores:
Pred. No.:              1.22e-220       Length:         653
Score:                  3190.50         Matches:        603
Percent Similarity:     97.90%          Conservative:   2
Best Local Similarity:  97.57%          Mismatches:     2
Query Match:            78.66%          Indels:         11
DB:                     5               Gaps:           1

US-09-914-958B-35 (1-2177) x AAO18264 (1-653)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                                :::|||:::|||   ||||||
Db        600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607

RESULT 4
AAE18679
ID   AAE18679 standard; protein; 653 AA.
XX
AC   AAE18679;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 protein.
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
```

```
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .21
FT                   /label= Signal_peptide
FT   Peptide         1. .19
FT                   /label= Signal_peptide
FT   Binding-site    12. .20
FT                   /note= "MHC class II DR4Dw4 binding motif"
FT   Protein         20. .653
FT                   /note= "Human mature SCIM-1 protein #1"
FT   Protein         22. .653
FT                   /note= "Human mature SCIM-1 protein #2"
FT   Binding-site    39. .47
FT                   /note= "MHC class II DR4Dw4 binding motif"
FT   Binding-site    149. .157
FT                   /note= "MHC class II DR4Dw4 binding motif"
FT   Binding-site    263. .265
FT                   /note= "Integrin-binding RGD motif"
FT   Binding-site    323. .331
FT                   /note= "MHC class II DR4Dw4 binding motif"
FT   Binding-site    551. .598
FT                   /note= "Calcium-binding EGF-like domain"
FT   Region          614. .635
FT                   /note= "Putative hydrophobic region"
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   N-PSDB; AAD29777.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 6; Page 30-32; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 protein
XX
SQ   Sequence 653 AA;
```

```
Alignment Scores:
Pred. No.:              1.22e-220       Length:         653
Score:                  3190.50         Matches:        603
Percent Similarity:     97.90%          Conservative:   2
Best Local Similarity:  97.57%          Mismatches:     2
Query Match:            78.66%          Indels:         11
DB:                     5               Gaps:           1

US-09-914-958B-35 (1-2177) x AAE18679 (1-653)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320
```

```
Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                      :::|||:::|||   ||||||
Db        600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607
```

RESULT 5
ABR48171

```
ID   ABR48171 standard; protein; 653 AA.
XX
AC   ABR48171;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Human bladder cancer associated protein sequence SEQ ID NO:58.
XX
KW   Human; bladder cancer; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2003003906-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US021338.
XX
PR   03-JUL-2001; 2001US-0302814P.
PR   03-AUG-2001; 2001US-0310099P.
PR   08-NOV-2001; 2001US-0343705P.
PR   13-NOV-2001; 2001US-0350666P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack DH,  Aziz N;
XX
DR   WPI; 2003-201532/19.
DR   N-PSDB; ACC50982.
XX
PT   Detecting a bladder cancer-associated transcript in a cell from a
PT   patient, comprises contacting a biological sample from the patient with a
PT   bladder cancer-associated polynucleotide or antibody.
XX
PS   Claim 10; Page 249; 307pp; English.
XX
CC   The present invention describes a method for detecting a bladder cancer-
CC   associated transcript in a cell from a patient. The method comprises
CC   contacting a biological sample from the patient with a polynucleotide
CC   that selectively hybridises to a sequence that is 80 % identical to a
CC   table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC   encode the human bladder cancer-associated proteins given in ABR48146 to
CC   ABR48242). Bladder cancer-associated sequences from the present invention
CC   have cytostatic activities, and can be used in antisense gene therapy and
CC   in vaccine production. The method can be used for detecting a bladder
CC   cancer-associated transcript in a cell from a patient. The method is
CC   useful in diagnosing or treating bladder cancer and in screening for
CC   compounds that modulate bladder cancer, such as hormones or antibodies.
CC   The nucleic acid molecules from the present invention may be used in
CC   various screening and diagnostic methods, and for gene therapy, vaccine
CC   and/or antisense/inhibition applications
XX
SQ   Sequence 653 AA;
```

Alignment Scores:

Pred. No.:	1.22e-220	Length:	653
Score:	3190.50	Matches:	603
Percent Similarity:	97.90%	Conservative:	2
Best Local Similarity:	97.57%	Mismatches:	2
Query Match:	78.66%	Indels:	11
DB:	6	Gaps:	1

US-09-914-958B-35 (1-2177) x ABR48171 (1-653)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy      1707 CTCTACCCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy      1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                                 :::|||:::|||   ||||||
Db       600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607

RESULT 6
ABU56636
ID   ABU56636 standard; protein; 653 AA.
XX
AC   ABU56636;
XX
DT   02-APR-2003  (first entry)
XX
DE   Lung cancer-associated polypeptide #229.
XX
KW   Lung cancer-associated polypeptide; cytostatic; emphysema;
KW   antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW   small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW   chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW   interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS   Unidentified.
XX
PN   WO200286443-A2.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US012476.
XX
PR   18-APR-2001; 2001US-0284770P.
PR   10-MAY-2001; 2001US-0290492P.
PR   09-NOV-2001; 2001US-0339245P.
PR   13-NOV-2001; 2001US-0350666P.
PR   29-NOV-2001; 2001US-0334370P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Aziz N,  Murray R;
XX
DR   WPI; 2003-093161/08.
DR   N-PSDB; ABX76365.
XX
PT   Detecting a lung cancer-associated transcript in a cell from a patient
PT   for treating lung cancer, by contacting a biological sample from the
```

PT patient with a polynucleotide that exhibits increased or decreased
PT expression in lung cancer.
XX
PS Claim 27; Page 366-367; 453pp; English.
XX
CC The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated polypeptide, for
CC inhibiting proliferation of a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer, such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
CC invention
XX
SQ Sequence 653 AA;

Alignment Scores:

Pred. No.:	1.22e-220	Length:	653
Score:	3190.50	Matches:	603
Percent Similarity:	97.90%	Conservative:	2
Best Local Similarity:	97.57%	Mismatches:	2
Query Match:	78.66%	Indels:	11
DB:	6	Gaps:	1

US-09-914-958B-35 (1-2177) x ABU56636 (1-653)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                              :::|||:::|||   ||||||
Db        600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607
```

RESULT 7
AAE18680
ID AAE18680 standard; protein; 637 AA.
XX
AC AAE18680;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 splice variant A protein.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy.
XX
OS Homo sapiens.
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR N-PSDB; AAD29791.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 6; Page 38-40; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for
CC induction of immunological tolerance to an autoantigen in patients
CC suffering from autoimmune disorders, more specifically rheumatoid
CC arthritis. The inflammatory disorders treated include diseases like
CC Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC multiple sclerosis and diabetes. Polypeptides of the invention are also
CC useful in a diagnostic method for the detection of activated autoreactive
CC T cells. The present sequence is human SCIM-1 splice variant A protein
XX

SQ Sequence 637 AA;

Alignment Scores:

Pred. No.:	5.37e-220	Length:	637
Score:	3181.50	Matches:	602
Percent Similarity:	97.89%	Conservative:	2
Best Local Similarity:	97.57%	Mismatches:	6
Query Match:	78.44%	Indels:	7
DB:	5	Gaps:	1

US-09-914-958B-35 (1-2177) x AAE18680 (1-637)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG------ 1940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGln 600

Qy       1941 ----------------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976
                             |||:::   |||         :::|||      |||
Db        601 ValAlaPheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 617

RESULT 8
AAE18682
ID   AAE18682 standard; protein; 612 AA.
XX
AC   AAE18682;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 splice variant C protein.
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy.
XX
OS   Homo sapiens.
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   N-PSDB; AAD29793.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 6; Page 43-45; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 splice variant C protein
XX
SQ   Sequence 612 AA;

Alignment Scores:
Pred. No.:                8.01e-220       Length:          612
Score:                    3179.00         Matches:         598
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              78.38%          Indels:          0
DB:                       5               Gaps:            0

US-09-914-958B-35 (1-2177) x AAE18682 (1-612)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
```

```
Db        61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db        81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db       101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db       121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG 1940
Db       581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysVal 598
```

```
RESULT 9
AAO18267
ID   AAO18267 standard; protein; 636 AA.
XX
AC   AAO18267;
XX
DT   26-SEP-2002  (first entry)
XX
DE   Human chondrocyte specific CEP-68 protein SEQ ID NO:25.
XX
KW   Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW   extracellular matrix protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 377
FT                   /note= "encoded by GGA"
FT   Misc-difference 433
FT                   /note= "encoded by TTC"
FT   Misc-difference 541
FT                   /note= "encoded by TTC"
FT   Misc-difference 550
FT                   /note= "encoded by ATG"
FT   Misc-difference 557
FT                   /note= "encoded by CAGTTC"
XX
PN   WO200253709-A1.
XX
PD   11-JUL-2002.
XX
PF   24-DEC-2001; 2001WO-EP015307.
XX
PR   05-JAN-2001; 2001DE-01000305.
XX
PA   (CYTO-) CYTONET GMBH & CO KG.
XX
PI   Richter W,  Steck E;
XX
```

DR WPI; 2002-528856/56.
DR N-PSDB; AAL47936.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 4; Page 98-100; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is an alternatively spliced version of the CEP-68
CC protein
XX
SQ Sequence 636 AA;

Alignment Scores:

Pred. No.:	8.5e-217	Length:	636
Score:	3137.00	Matches:	597
Percent Similarity:	97.08%	Conservative:	2
Best Local Similarity:	96.76%	Mismatches:	10
Query Match:	77.34%	Indels:	8
DB:	5	Gaps:	2

US-09-914-958B-35 (1-2177) x AAO18267 (1-636)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGluAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyValAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ValSerGlnGlnGluAsnGlyHisCysAsnAspThrAsnGluCysIle---PheProPhe 559

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 579

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG------ 1940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        580 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGln 599

Qy       1941 ----------------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976
                              |||:::   |||         :::|||      |||
Db        600 ValAlaPheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 616

RESULT 10
ABR48172
ID   ABR48172 standard; protein; 1597 AA.
XX
AC   ABR48172;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Human bladder cancer associated protein sequence SEQ ID NO:60.
XX
KW   Human; bladder cancer; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2003003906-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US021338.
XX
PR   03-JUL-2001; 2001US-0302814P.
PR   03-AUG-2001; 2001US-0310099P.
PR   08-NOV-2001; 2001US-0343705P.
PR   13-NOV-2001; 2001US-0350666P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack DH,  Aziz N;
XX
DR   WPI; 2003-201532/19.
DR   N-PSDB; ACC50983.
XX
PT   Detecting a bladder cancer-associated transcript in a cell from a
PT   patient, comprises contacting a biological sample from the patient with a
PT   bladder cancer-associated polynucleotide or antibody.
XX
PS   Claim 10; Page 250; 307pp; English.
XX
CC   The present invention describes a method for detecting a bladder cancer-
CC   associated transcript in a cell from a patient. The method comprises
CC   contacting a biological sample from the patient with a polynucleotide
CC   that selectively hybridises to a sequence that is 80 % identical to a
CC   table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC   encode the human bladder cancer-associated proteins given in ABR48146 to
CC   ABR48242). Bladder cancer-associated sequences from the present invention
CC   have cytostatic activities, and can be used in antisense gene therapy and
CC   in vaccine production. The method can be used for detecting a bladder
CC   cancer-associated transcript in a cell from a patient. The method is
CC   useful in diagnosing or treating bladder cancer and in screening for
CC   compounds that modulate bladder cancer, such as hormones or antibodies.
CC   The nucleic acid molecules from the present invention may be used in
CC   various screening and diagnostic methods, and for gene therapy, vaccine
CC   and/or antisense/inhibition applications
XX
SQ   Sequence 1597 AA;

Alignment Scores:
Pred. No.:              3.95e-159      Length:          1597
Score:                  2337.00        Matches:         566
Percent Similarity:     35.18%         Conservative:    6
Best Local Similarity:  34.81%         Mismatches:      29
Query Match:            57.62%         Indels:          1025
DB:                     6              Gaps:            11

US-09-914-958B-35 (1-2177) x ABR48172 (1-1597)

Qy       145 GCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGT 204
             ||||||||||||                    |||||||||:::         |||   |||
Db         2 AlaCysProGlyGlyLeu---ProAlaArgCysSerGlyTrpMet---GlyLeuGlyGly 19

Qy       205 CCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATG 264
             ||||||||   :::|||   |||
Db        20 ProSerGlySerSerProAlaSer---------------------------------- 27

Qy       265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
                 ||||||
Db        28 ---------ProPro---------------------------------------------- 29

Qy       325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
                 ||||||    |||  ||||||||||||||||||||||||||||||||||||||||||
Db        30 ------HisSerSerSerArg--TyrAsnGlyProAsnLeuValLeuLysTyrAspArgA 47

Qy       385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        47 laGlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuA 67

Qy       445 GGGACCGGCAGGCGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        67 rgAspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgG 87

Qy       505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTC------------------------- 539
             |||||||||||||||||||||||||||||||||
Db        87 luGluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyHisSerSerSerAlaGlnV 107

Qy       539 ------------------------------------------------------------ 539

Db       107 alProSerGlyLeuHisArgAsnArgProValLeuLysProProProThrThrProAlaG 127

Qy       539 ------------------------------------------------------------ 539

Db       127 lyLeuLeuGlyLeuProProLeuSerGlyArgAspPheSerSerSerLeuGlyGlnAlaS 147

Qy       539 ------------------------------------------------------------ 539

Db       147 erProAspSerArgGlnGlyGluArgValProValProCysCysArgGlyGlyLeuArgP 167

Qy       540 ------------------------------------TCGGGGGTGGCCACGTACACCG 561
                                                 ||||||||||||||||||||||
Db       167 roThrHisGluProGluProPheLeuLeuArgProLysSerGlyValAlaThrTyrThrA 187

Qy       562 ACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACG 621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       187 spLysLeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnV 207

Qy       622 TGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGG 681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       207 alAlaArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysG 227

Qy       682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATG 741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       227 lySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspA 247

Qy       742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       247 laLeuIleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgA 267

Qy       802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACA------------------------- 836
             |||||||||||||||||||||||||||||||||||
Db       267 spValAlaAlaGluAlaGlyValSerLysTyrThrGluGlyPheSerHisThrAlaSerP 287

Qy       836 ------------------------------------------------------------ 836

Db       287 roSerIleGlyGluIleSerGlyArgThrGluGluArgGluGlyGlyAspProGluGluA 307

Qy       836 ------------------------------------------------------------ 836

Db       307 laAspGluGluHisSerGlyAspGlySerThrSerGlnLeuCysArgLeuGlyTrpLysA 327

Qy       836 ------------------------------------------------------------ 836
```

```
Db      327 spGlyGlnPheLysGluGluAlaAlaAlaLeuValGluGluGlnArgGluAlaGlyAlaA 347
Qy      836 ------------------------------------------------------------ 836
Db      347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAlaA 367
Qy      836 ------------------------------------------------------------ 836
Db      367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387
Qy      836 ------------------------------------------------------------ 836
Db      387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrGlnLeuM 407
Qy      836 ------------------------------------------------------------ 836
Db      407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427
Qy      836 ------------------------------------------------------------ 836
Db      427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447
Qy      836 ------------------------------------------------------------ 836
Db      447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467
Qy      836 ------------------------------------------------------------ 836
Db      467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487
Qy      836 ------------------------------------------------------------ 836
Db      487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProProIleL 507
Qy      836 ------------------------------------------------------------ 836
Db      507 euGlnArgThrAspGlyAspProGlyArgArgArgAspSerProLysValThrGlnGluC 527
Qy      836 ------------------------------------------------------------ 836
Db      527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547
Qy      836 ------------------------------------------------------------ 836
Db      547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567
Qy      836 ------------------------------------------------------------ 836
Db      567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587
Qy      836 ------------------------------------------------------------ 836
Db      587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607
Qy      836 ------------------------------------------------------------ 836
Db      607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627
Qy      836 ------------------------------------------------------------ 836
Db      627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647
Qy      836 ------------------------------------------------------------ 836
Db      647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlyS 667
Qy      837 --------------GGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
                            ||||||||||||||||||||||||||||||||||||||||||||
Db      667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS 687
Qy      883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707
Qy      943 GCACCTTTGTGGACGCTGCGGCCAGTGCT---------------------------------- 971
                                        ||||||||||||||||||||||||||||||
Db      707 lyThrPheValAspAlaAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727
Qy      971 ------------------------------------------------------------ 971
Db      727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747
Qy      971 ------------------------------------------------------------ 971
Db      747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisHisGlyL 767
Qy      971 ------------------------------------------------------------ 971
Db      767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787
Qy      971 ------------------------------------------------------------ 971
Db      787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807
Qy      971 ------------------------------------------------------------ 971
Db      807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827
Qy      971 ------------------------------------------------------------ 971
Db      827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847
Qy      972 -----GGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACC 1026
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867
Qy     1027 GTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGC 1086
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887
Qy     1087 AAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGC 1146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      887 lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP 907
Qy     1147 CCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCT 1206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927
Qy     1207 TCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGC---------- 1256
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947
Qy     1256 ------------------------------------------------------------ 1256
Db      947 euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967
Qy     1256 ------------------------------------------------------------ 1256
Db      967 rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL 987
Qy     1256 ------------------------------------------------------------ 1256
Db      987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007
Qy     1256 ------------------------------------------------------------ 1256
Db     1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027
Qy     1256 ------------------------------------------------------------ 1256
Db     1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047
Qy     1256 ------------------------------------------------------------ 1256
Db     1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067
```

```
Qy      1256 ------------------------------------------------------------ 1256
Db      1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087
Qy      1256 ------------------------------------------------------------ 1256
Db      1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107
Qy      1256 ------------------------------------------------------------ 1256
Db      1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127
Qy      1256 ------------------------------------------------------------ 1256
Db      1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147
Qy      1256 ------------------------------------------------------------ 1256
Db      1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167
Qy      1256 ------------------------------------------------------------ 1256
Db      1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187
Qy      1256 ------------------------------------------------------------ 1256
Db      1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207
Qy      1256 ------------------------------------------------------------ 1256
Db      1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227
Qy      1256 ------------------------------------------------------------ 1256
Db      1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247
Qy      1256 ------------------------------------------------------------ 1256
Db      1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267
Qy      1256 ------------------------------------------------------------ 1256
Db      1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287
Qy      1257 --------------------------------GTCATCCGTAGAGAGCACGGAGACCCCC 1284
                                               |||||||||||||||||||||||||||
Db      1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307
Qy      1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327
Qy      1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347
Qy      1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367
Qy      1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1367 alValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrL 1387
Qy      1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1387 ysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetG 1407
Qy      1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427

Qy      1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447
Qy      1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467
Qy      1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487
Qy      1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507
Qy      1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGC- 1943
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527
Qy      1944 --------------------------TGGTGGAGCCCTGTGTTGAAG------------- 1964
                                       |||   |||     |||
Db      1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547
Qy      1965 -----------ATAGTGACACCACAAG------------------------------- 1980
                        |||   ||||||
Db      1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566
Qy      1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
             |||   ||||||  :::     :::  :::     |||:::     |||||||||
Db      1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586

Qy      2041 CCTGTTGATCAG 2052
             |||   |||
Db      1587 ProGluAlaGln 1590
```

RESULT 11
ABU56637
ID ABU56637 standard; protein; 1597 AA.
XX
AC ABU56637;
XX
DT 02-APR-2003 (first entry)
XX
DE Lung cancer-associated polypeptide #230.
XX
KW Lung cancer-associated polypeptide; cytostatic; emphysema;
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS Unidentified.
XX
PN WO200286443-A2.
XX
PD 31-OCT-2002.
XX
PF 18-APR-2002; 2002WO-US012476.
XX
PR 18-APR-2001; 2001US-0284770P.
PR 10-MAY-2001; 2001US-0290492P.
PR 09-NOV-2001; 2001US-0339245P.
PR 13-NOV-2001; 2001US-0350666P.
PR 29-NOV-2001; 2001US-0334370P.
PR 12-APR-2002; 2002US-0372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Aziz N, Murray R;
XX
DR WPI; 2003-093161/08.
DR N-PSDB; ABX76366.

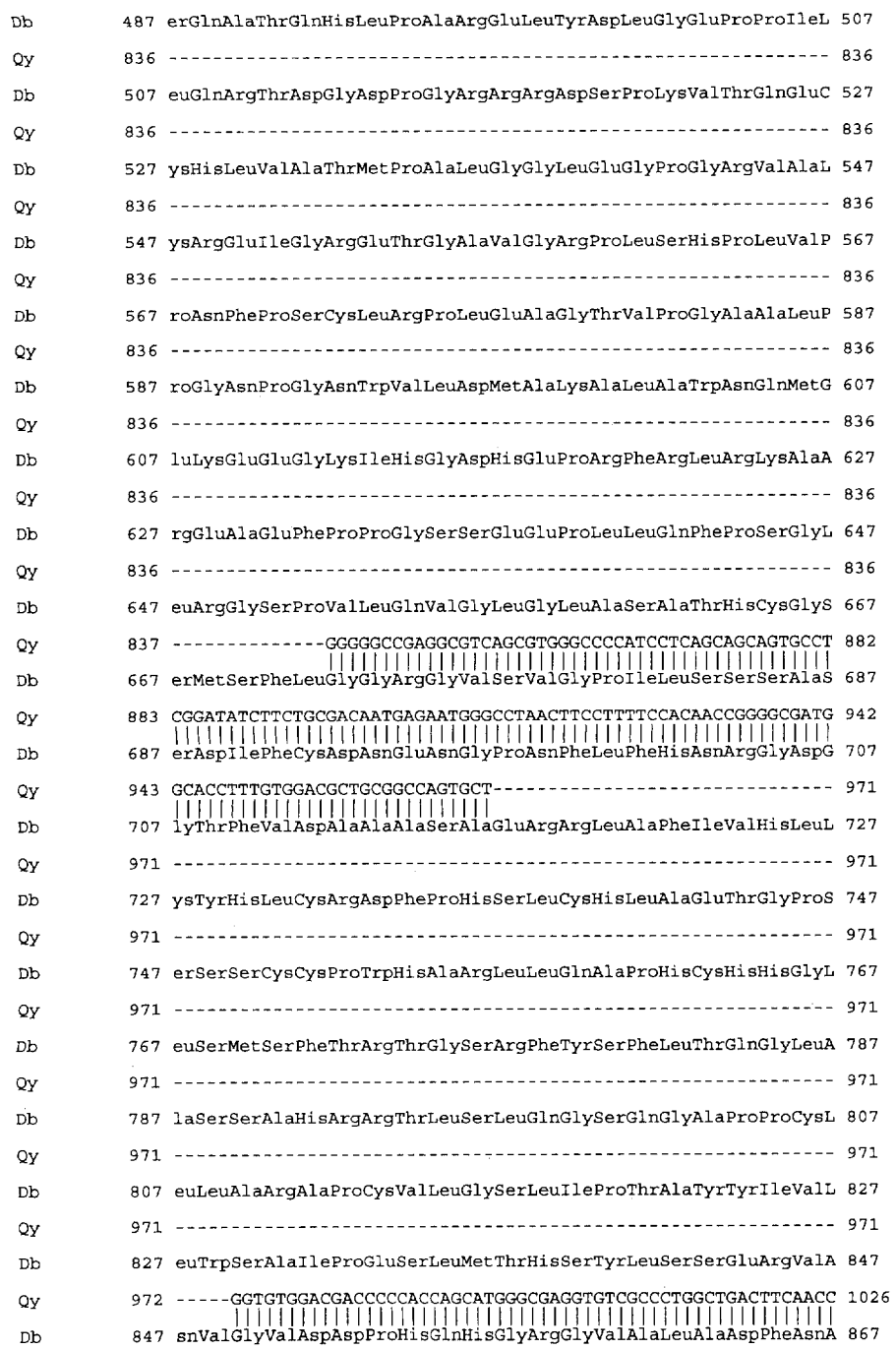
XX
PT Detecting a lung cancer-associated transcript in a cell from a patient
PT for treating lung cancer, by contacting a biological sample from the
PT patient with a polynucleotide that exhibits increased or decreased
PT expression in lung cancer.
XX
PS Claim 27; Page 368; 453pp; English.
XX
CC The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated polypeptide, for
CC inhibiting proliferation of a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer, such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
CC invention
XX
SQ Sequence 1597 AA;

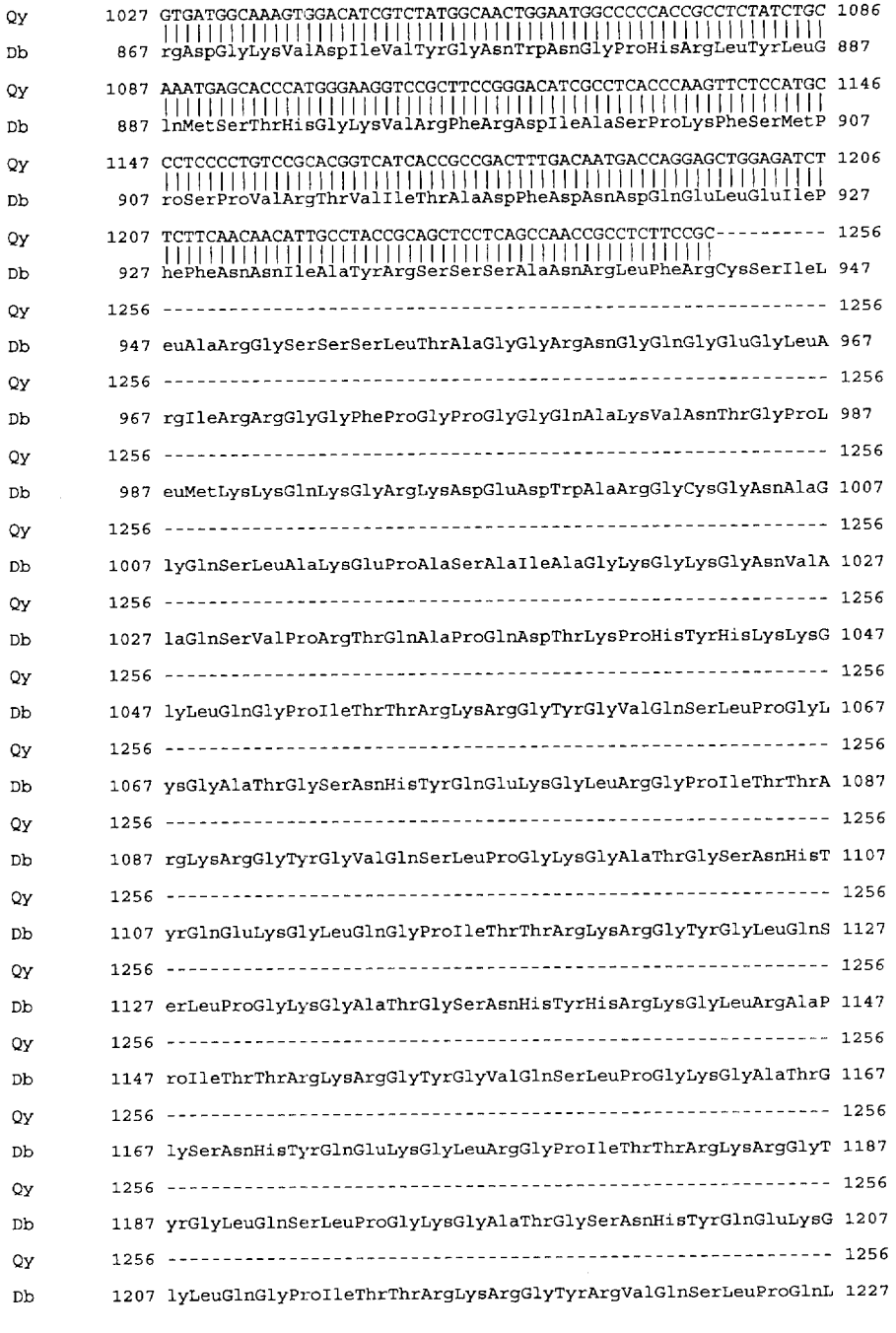
Alignment Scores:

Pred. No.:	3.95e-159	Length:	1597
Score:	2337.00	Matches:	566
Percent Similarity:	35.18%	Conservative:	6
Best Local Similarity:	34.81%	Mismatches:	29
Query Match:	57.62%	Indels:	1025
DB:	6	Gaps:	11

US-09-914-958B-35 (1-2177) x ABU56637 (1-1597)

```
Qy        145 GCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGT 204
              ||||||||||||                          |||||||||:::         |||   |||
Db          2 AlaCysProGlyGlyLeu---ProAlaArgCysSerGlyTrpMet---GlyLeuGlyGly 19

Qy        205 CCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATG 264
              |||||||||   :::|||   |||
Db         20 ProSerGlySerSerProAlaSer---------------------------------- 27

Qy        265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
                      ||||||
Db         28 ---------ProPro------------------------------------------- 29

Qy        325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
                      ||||||   |||  ||||||||||||||||||||||||||||||||||||||
Db         30 ------HisSerSerSerArg--TyrAsnGlyProAsnLeuValLeuLysTyrAspArgA 47

Qy        385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 laGlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuA 67

Qy        445 GGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 rgAspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgG 87

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTC---------------------------- 539
              ||||||||||||||||||||||||||||||||
Db         87 luGluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyHisSerSerSerAlaGlnV 107

Qy        539 ------------------------------------------------------------ 539

Db        107 alProSerGlyLeuHisArgAsnArgProValLeuLysProProProThrThrProAlaG 127

Qy        539 ------------------------------------------------------------ 539

Db        127 lyLeuLeuGlyLeuProProLeuSerGlyArgAspPheSerSerSerLeuGlyGlnAlaS 147

Qy        539 ------------------------------------------------------------ 539

Db        147 erProAspSerArgGlnGlyGluArgValProValProCysCysArgGlyGlyLeuArgP 167

Qy        540 ------------------------------------TCGGGGGTGGCCACGTACACCG 561
                                                 ||||||||||||||||||||||
Db        167 roThrHisGluProGluProPheLeuLeuArgProLysSerGlyValAlaThrTyrThrA 187

Qy        562 ACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACG 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 spLysLeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnV 207

Qy        622 TGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGG 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 alAlaArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysG 227

Qy        682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATG 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 lySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspA 247

Qy        742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 laLeuIleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgA 267

Qy        802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACA------------------------- 836
              ||||||||||||||||||||||||||||||||||
Db        267 spValAlaAlaGluAlaGlyValSerLysTyrThrGluGlyPheSerHisThrAlaSerP 287

Qy        836 ------------------------------------------------------------ 836

Db        287 roSerIleGlyGluIleSerGlyArgThrGluGluArgGluGlyGlyAspProGluGluA 307

Qy        836 ------------------------------------------------------------ 836

Db        307 laAspGluGluHisSerGlyAspGlySerThrSerGlnLeuCysArgLeuGlyTrpLysA 327

Qy        836 ------------------------------------------------------------ 836

Db        327 spGlyGlnPheLysGluGluAlaAlaAlaLeuValGluGluGlnArgGluAlaGlyAlaA 347

Qy        836 ------------------------------------------------------------ 836

Db        347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAlaA 367

Qy        836 ------------------------------------------------------------ 836

Db        367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387

Qy        836 ------------------------------------------------------------ 836

Db        387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrAlaLeuM 407

Qy        836 ------------------------------------------------------------ 836

Db        407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427

Qy        836 ------------------------------------------------------------ 836

Db        427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447

Qy        836 ------------------------------------------------------------ 836

Db        447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467

Qy        836 ------------------------------------------------------------ 836

Db        467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487

Qy        836 ------------------------------------------------------------ 836
```


Db 487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProProIleL 507
Qy 836 ---------- 836
Db 507 euGlnArgThrAspGlyAspProGlyArgArgArgAspSerProLysValThrGlnGluC 527
Qy 836 ---------- 836
Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547
Qy 836 ---------- 836
Db 547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567
Qy 836 ---------- 836
Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587
Qy 836 ---------- 836
Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607
Qy 836 ---------- 836
Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627
Qy 836 ---------- 836
Db 627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647
Qy 836 ---------- 836
Db 647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlyS 667
Qy 837 ----------GGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS 687
Qy 883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707
Qy 943 GCACCTTTGTGGACGCTGCGGCCAGTGCT---------- 971
Db 707 lyThrPheValAspAlaAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727
Qy 971 ---------- 971
Db 727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747
Qy 971 ---------- 971
Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisHisGlyL 767
Qy 971 ---------- 971
Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787
Qy 971 ---------- 971
Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807
Qy 971 ---------- 971
Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827
Qy 971 ---------- 971
Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847
Qy 972 -----GGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACC 1026
Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867


Qy 1027 GTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGC 1086
Db 867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887
Qy 1087 AAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGC 1146
Db 887 lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP 907
Qy 1147 CCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCT 1206
Db 907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927
Qy 1207 TCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGC---------- 1256
Db 927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947
Qy 1256 ---------- 1256
Db 947 euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967
Qy 1256 ---------- 1256
Db 967 rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL 987
Qy 1256 ---------- 1256
Db 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007
Qy 1256 ---------- 1256
Db 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027
Qy 1256 ---------- 1256
Db 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047
Qy 1256 ---------- 1256
Db 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067
Qy 1256 ---------- 1256
Db 1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087
Qy 1256 ---------- 1256
Db 1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107
Qy 1256 ---------- 1256
Db 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127
Qy 1256 ---------- 1256
Db 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147
Qy 1256 ---------- 1256
Db 1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167
Qy 1256 ---------- 1256
Db 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187
Qy 1256 ---------- 1256
Db 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207
Qy 1256 ---------- 1256
Db 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227

```
Qy       1256 ------------------------------------------------------------ 1256

Db       1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247

Qy       1256 ------------------------------------------------------------ 1256

Db       1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267

Qy       1256 ------------------------------------------------------------ 1256

Db       1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287

Qy       1257 --------------------------------GTCATCCGTAGAGAGCACGGAGACCCCC 1284
                                               ||||||||||||||||||||||||||
Db       1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307

Qy       1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327

Qy       1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347

Qy       1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367

Qy       1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 alValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrL 1387

Qy       1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 ysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetG 1407

Qy       1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427

Qy       1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447

Qy       1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467

Qy       1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487

Qy       1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507

Qy       1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGC- 1943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db       1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527

Qy       1944 --------------------------TGGTGGAGCCCTGTGTTGAAG------------- 1964
                                         |||  |||     |||
Db       1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547

Qy       1965 -----------ATAGTGACACCACAAG------------------------------- 1980
                         |||  ||||||
Db       1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566

Qy       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
              |||  ||||||  :::     :::  :::     ||:::     ||||||||
Db       1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586

Qy       2041 CCTGTTGATCAG 2052
```

```
                |||     |||
Db       1587 ProGluAlaGln 1590
```

RESULT 12
AAM39983
ID AAM39983 standard; protein; 435 AA.
XX
AC AAM39983;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3128.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59139.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 4; SEQ ID NO 3128; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 435 AA;

Alignment Scores:

Pred. No.:	4.69e-144	Length:	435
Score:	2124.50	Matches:	400

```
Percent Similarity:     96.87%          Conservative:  2
Best Local Similarity:  96.39%          Mismatches:    6
Query Match:            52.38%          Indels:        7
DB:                     4               Gaps:          1

US-09-914-958B-35 (1-2177) x AAM39983 (1-435)

Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSer 340

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnProPheValCys 360

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG---------- 1940
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGlnValAla 400

Qy       1941 ---------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976
                        |||:::  |||        :::|||     |||
Db        401 PheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 415

RESULT 13
AAB92693
ID   AAB92693 standard; prctein; 418 AA.
XX
AC   AAB92693;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11083.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 11083; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
```

```
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 418 AA;

Alignment Scores:
Pred. No.:                8.99e-133      Length:          418
Score:                    1967.50        Matches:         385
Percent Similarity:       89.81%         Conservative:    3
Best Local Similarity:    89.12%         Mismatches:      10
Query Match:              48.51%         Indels:          34
DB:                       4              Gaps:            2

US-09-914-958B-35 (1-2177) x AAB92693 (1-418)

Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu---------------- 334

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
                                                ||||||||||||||||||||||||||||||
Db        335 ------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
                                                                  |||   ::
Db        384 ------------------------------------------------LeuSerAlaSe 387

Qy       2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
              :|||         :::||||||      |||:::
Db        387 rHisArgAlaProAlaProProProProProLeu 398

RESULT 14
ABR48170
ID   ABR48170 standard; protein; 418 AA.
XX
AC   ABR48170;
XX
DT   12-JUN-2003  (first entry)
XX
DE   Human bladder cancer associated protein sequence SEQ ID NO:56.
XX
KW   Human; bladder cancer; cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2003003906-A2.
XX
PD   16-JAN-2003.
XX
PF   03-JUL-2002; 2002WO-US021338.
XX
PR   03-JUL-2001; 2001US-0302814P.
PR   03-AUG-2001; 2001US-0310099P.
PR   08-NOV-2001; 2001US-0343705P.
PR   13-NOV-2001; 2001US-0350666P.
PR   12-APR-2002; 2002US-0372246P.
XX
PA   (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI   Mack DH,  Aziz N;
XX
DR   WPI; 2003-201532/19.
DR   N-PSDB; ACC50981.
XX
PT   Detecting a bladder cancer-associated transcript in a cell from a
PT   patient, comprises contacting a biological sample from the patient with a
PT   bladder cancer-associated polynucleotide or antibody.
XX
PS   Claim 10; Page 248; 307pp; English.
```

```
XX
CC   The present invention describes a method for detecting a bladder cancer-
CC   associated transcript in a cell from a patient. The method comprises
CC   contacting a biological sample from the patient with a polynucleotide
CC   that selectively hybridises to a sequence that is 80 % identical to a
CC   table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC   encode the human bladder cancer-associated proteins given in ABR48146 to
CC   ABR48242). Bladder cancer-associated sequences from the present invention
CC   have cytostatic activities, and can be used in antisense gene therapy and
CC   in vaccine production. The method can be used for detecting a bladder
CC   cancer-associated transcript in a cell from a patient. The method is
CC   useful in diagnosing or treating bladder cancer and in screening for
CC   compounds that modulate bladder cancer, such as hormones or antibodies.
CC   The nucleic acid molecules from the present invention may be used in
CC   various screening and diagnostic methods, and for gene therapy, vaccine
CC   and/or antisense/inhibition applications
XX
SQ   Sequence 418 AA;

Alignment Scores:
Pred. No.:                8.99e-133      Length:          418
Score:                    1967.50        Matches:         385
Percent Similarity:       89.81%         Conservative:    3
Best Local Similarity:    89.12%         Mismatches:      10
Query Match:              48.51%         Indels:          34
DB:                       6              Gaps:            2

US-09-914-958B-35 (1-2177) x ABR48170 (1-418)

Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu---------------- 334

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
                                               ||||||||||||||||||||||||||||||
Db        335 -------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
                                                                  ||| ::
Db        384 ------------------------------------------------LeuSerAlaSe 387

Qy       2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
              :|||       :::||||||      |||:::
Db        387 rHisArgAlaProAlaProProProProProLeu 398

RESULT 15
ABU56635
ID   ABU56635 standard; protein; 418 AA.
XX
AC   ABU56635;
XX
DT   02-APR-2003  (first entry)
XX
DE   Lung cancer-associated polypeptide #228.
XX
KW   Lung cancer-associated polypeptide; cytostatic; emphysema;
KW   antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW   small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW   chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW   interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS   Unidentified.
XX
PN   WO200286443-A2.
XX
PD   31-OCT-2002.
XX
PF   18-APR-2002; 2002WO-US012476.
XX
PR   18-APR-2001; 2001US-0284770P.
PR   10-MAY-2001; 2001US-0290492P.
PR   09-NOV-2001; 2001US-0339245P.
PR   13-NOV-2001; 2001US-0350666P.
PR   29-NOV-2001; 2001US-0334370P.
```

PR 12-APR-2002; 2002US-0372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Aziz N, Murray R;
XX
DR WPI; 2003-093161/08.
DR N-PSDB; ABX76364.
XX
PT Detecting a lung cancer-associated transcript in a cell from a patient
PT for treating lung cancer, by contacting a biological sample from the
PT patient with a polynucleotide that exhibits increased or decreased
PT expression in lung cancer.
XX
PS Claim 27; Page 366; 453pp; English.
XX
CC The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated polypeptide, for
CC inhibiting proliferation of a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer, such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
CC invention
XX
SQ Sequence 418 AA;

Alignment Scores:

Pred. No.:	8.99e-133	Length:	418
Score:	1967.50	Matches:	385
Percent Similarity:	89.81%	Conservative:	3
Best Local Similarity:	89.12%	Mismatches:	10
Query Match:	48.51%	Indels:	34
DB:	6	Gaps:	2

US-09-914-958B-35 (1-2177) x ABU56635 (1-418)

```
Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu---------------- 334

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
                                              ||||||||||||||||||||||||||||||
Db        335 ------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
                                                                  |||   ::
Db        384 ----------------------------------------------LeuSerAlaSe 387

Qy       2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
              :|||       :::|||||     |||:::
Db        387 rHisArgAlaProAlaProProProProProLeu 398
```

Search completed: August 3, 2004, 19:43:21
Job time : 141 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:39:15 ; Search time 24.5 Seconds
(without alignments)
9174.673 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161652_29729/app_query.fasta_1.2375
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914958_@CGN_1_1_27_@runat_03082004_161652_29729 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	247	6.0	1093	4	US-09-252-991A-21827	Sequence 21827, A
c	2	240	5.8	1706	4	US-09-252-991A-31760	Sequence 31760, A
c	3	239.5	5.8	467	4	US-09-252-991A-18296	Sequence 18296, A
c	4	234	5.7	681	4	US-09-252-991A-24567	Sequence 24567, A
	5	233.5	5.8	1093	4	US-09-252-991A-21827	Sequence 21827, A
	6	217	5.4	663	4	US-09-252-991A-30843	Sequence 30843, A
	7	215.5	5.3	720	4	US-09-252-991A-21881	Sequence 21881, A
c	8	215	5.2	595	4	US-09-252-991A-18995	Sequence 18995, A
c	9	212.5	5.1	590	4	US-09-252-991A-19127	Sequence 19127, A
c	10	212.5	5.1	638	4	US-09-252-991A-27068	Sequence 27068, A
c	11	212	5.1	2294	4	US-09-252-991A-17231	Sequence 17231, A
	12	210.5	5.2	798	4	US-09-252-991A-23774	Sequence 23774, A
c	13	210	5.1	977	4	US-09-252-991A-16655	Sequence 16655, A
c	14	207	5.0	663	4	US-09-252-991A-30843	Sequence 30843, A
c	15	203.5	4.9	822	4	US-09-252-991A-21920	Sequence 21920, A
	16	202.5	5.0	878	4	US-09-556-706B-2	Sequence 2, Appli
	17	202.5	5.0	977	4	US-09-252-991A-16655	Sequence 16655, A
c	18	201	4.9	1418	4	US-09-252-991A-32367	Sequence 32367, A
c	19	199.5	4.8	561	4	US-09-252-991A-23080	Sequence 23080, A
c	20	198	4.8	618	4	US-09-252-991A-23696	Sequence 23696, A
c	21	198	4.8	1411	4	US-09-252-991A-28408	Sequence 28408, A
c	22	197.5	4.8	552	3	US-09-219-849-7	Sequence 7, Appli
	23	197.5	4.9	575	4	US-09-252-991A-20649	Sequence 20649, A
c	24	196.5	4.8	569	4	US-09-252-991A-24513	Sequence 24513, A
c	25	196.5	4.8	863	4	US-09-252-991A-26099	Sequence 26099, A
	26	196.5	4.8	907	3	US-08-783-774-2	Sequence 2, Appli
	27	196.5	4.8	907	4	US-09-328-599A-1	Sequence 1, Appli
	28	196.5	4.8	907	5	PCT-US95-04611A-19	Sequence 19, Appl
c	29	195.5	4.7	686	4	US-09-252-991A-20509	Sequence 20509, A
c	30	195	4.7	507	4	US-09-252-991A-23598	Sequence 23598, A
c	31	194.5	4.7	950	4	US-09-252-991A-25927	Sequence 25927, A
	32	194	4.8	987	4	US-09-252-991A-27483	Sequence 27483, A
c	33	193.5	4.7	537	4	US-09-252-991A-27024	Sequence 27024, A
c	34	193.5	4.7	720	4	US-09-252-991A-21881	Sequence 21881, A
c	35	192	4.6	660	4	US-09-252-991A-22842	Sequence 22842, A
c	36	191.5	4.6	550	4	US-09-252-991A-17682	Sequence 17682, A
	37	191.5	4.7	611	4	US-09-252-991A-17597	Sequence 17597, A
c	38	191.5	4.6	1185	3	US-09-041-886-23	Sequence 23, Appl
	39	191	4.7	610	4	US-09-252-991A-20409	Sequence 20409, A
c	40	190.5	4.6	654	4	US-09-252-991A-25801	Sequence 25801, A
c	41	190	4.6	556	4	US-09-252-991A-22670	Sequence 22670, A
c	42	189.5	4.6	611	4	US-09-252-991A-32402	Sequence 32402, A
c	43	189.5	4.6	960	3	US-09-219-849-6	Sequence 6, Appli
c	44	189	4.6	476	4	US-09-252-991A-27078	Sequence 27078, A
c	45	189	4.6	1002	4	US-09-252-991A-27980	Sequence 27980, A

ALIGNMENTS

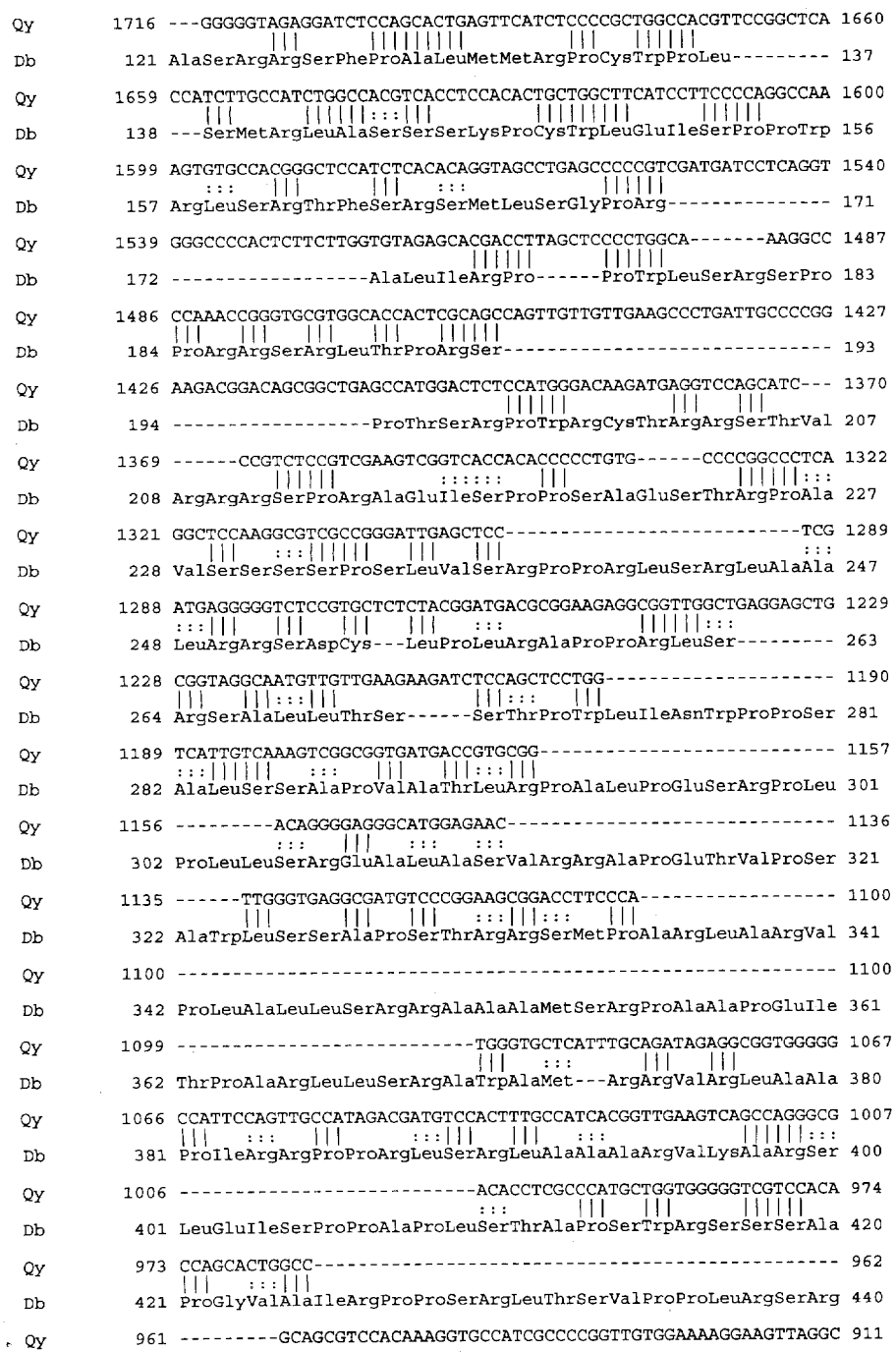
RESULT 1
US-09-252-991A-21827
; Sequence 21827, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21827
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21827

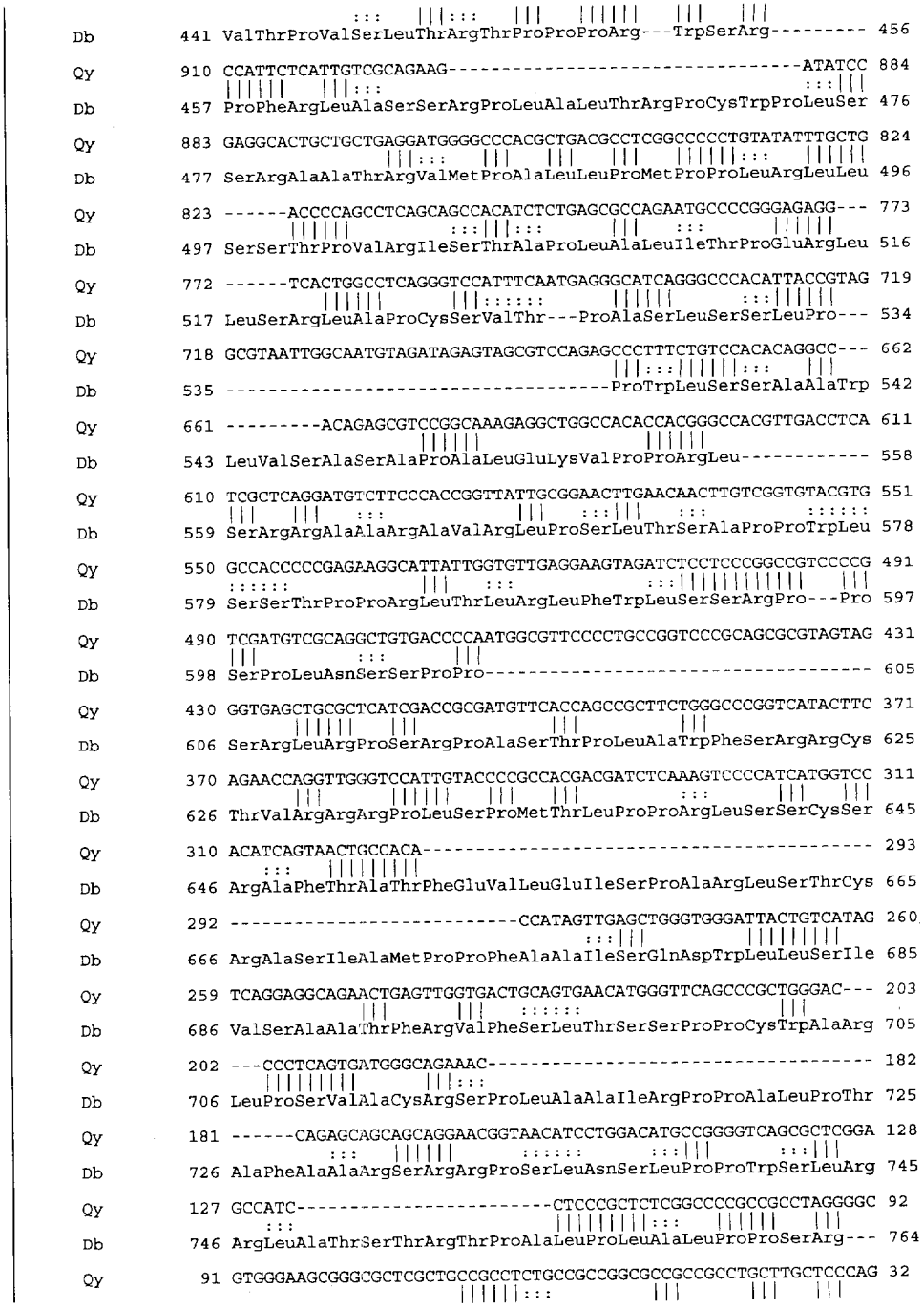
Alignment Scores:

Pred. No.:	1.26e-10	Length:	1093
Score:	247.00	Matches:	183
Percent Similarity:	32.64%	Conservative:	68
Best Local Similarity:	23.80%	Mismatches:	254
Query Match:	5.97%	Indels:	265
DB:	4	Gaps:	36

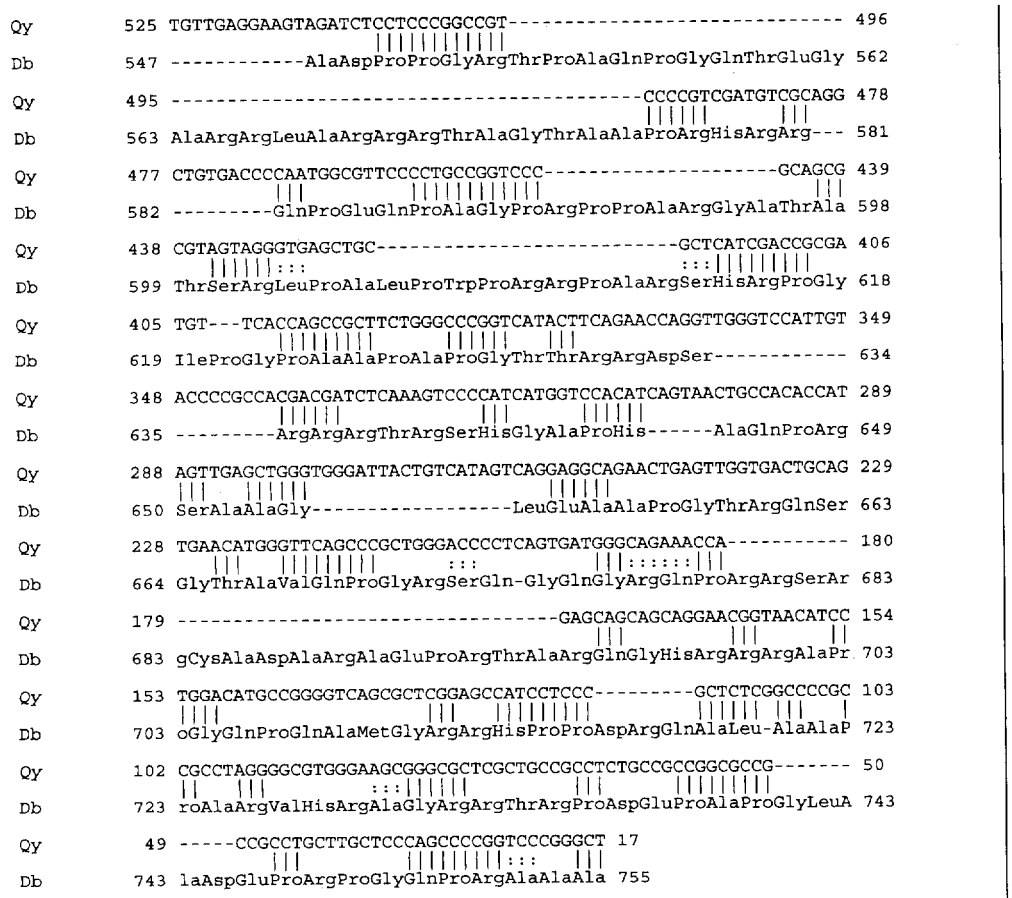
US-09-914-958B-35 (1-2177) x US-09-252-991A-21827 (1-1093)

```
Qy       1755 ACTCCAGTGGGGCTGGGT---CCTGAAGTGTGTCCTCATCCC------------------ 1717
                 :::|||   |||   |||   ||||||      |||   |||
Db        101 SerProSerGlyProGlyHisProGluSerArgProAlaProAlaLeuProSerAlaSer 120
```





```
Db        765 ---------LeuSerLysAlaProProValAlaLeuLysProAlaSerAlaThrSerThr 781

Qy         31 CCCCGGTCCCGGGCTGGCCTCGAGCCT 5
              |||          :::    |||   |||
Db        782 ProProArgLeuSerArgLeuAlaPro 790

RESULT 2
US-09-252-991A-31760
; Sequence 31760, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 31760
;   LENGTH: 1706
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31760

Alignment Scores:
Pred. No.:                5.24e-10      Length:          1706
Score:                    240.00        Matches:         190
Percent Similarity:       29.12%        Conservative:    36
Best Local Similarity:    24.48%        Mismatches:      219
Query Match:              5.81%         Indels:          331
DB:                       4             Gaps:            40

US-09-914-958B-35 (1-2177) x US-09-252-991A-31760 (1-1706)

Qy       1779 CCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTGGGTCCTGAAGTGTGTCCTCAT 1720
              ||||||   |||:::                  |||   |||   :::
Db        126 ProAlaAspArgLeuProGluAlaArgArgProGlySerGlyAlaGlnAla--------- 142

Qy       1719 CCCGGGGGTAGAGGA---TCTCCAGCACTGAGTTCATCTCCCCGC---------TGGCCA 1672
              ||||||||||||||    ||||||          :::   |||           |||
Db        143 ProGlyGlyArgGlyProGlyProAlaHisArgHisAlaGlnArgArgGlnProArgPro 162

Qy       1671 CGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCT 1612
              |||:::|||   |||                                        ||||||
Db        163 ArgThrGlyAspProGln-----------------------------------HisPro 170

Qy       1611 ------------------------------------TCCCCAGGCCAAAGTGTGCCA--- 1591
                                                  ||||||          |||
Db        171 ArgAlaAlaArgSerAlaValArgArgSerArgArgArgProGlyLeuGlyAlaProAla 190

Qy       1590 CGGGCTCCATCT----------------------CACACAGGTAGCCTGAGCCCCCGT 1555
              |||||||||                      ||||||   |||   |||
Db        191 ArgAlaProGlyGlyAlaTyrArgGlyAlaGlyArgArgThrGlyProLeuProProGlu 210

Qy       1554 CGATGATCCTCAGGTGGGCCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTCC----- 1500
                    |||      |||           |||||| :::|||   :::
Db        211 ProAlaProArgGlyArgGlyHis------------ArgAla-GluLeuProAlaProAl 226

Qy       1499 ------------------------------------------------CCTGGCAAAGGCCC 1486
                                                              ||||||      ||
Db        226 aArgArgThrAspArgAlaGlyGluArgLeuAlaAlaAlaGlyLeuProGlyProAlaPr 246

Qy       1485 CAAACCGGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGAAGCCCTGATTGCCCCGGA 1426
              |   ||||||          |||   ||||||
Db        246 oAlaProGlyGlyAlaGlyHisArgGlnPro---------------------------- 256

Qy       1425 AGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGATGAGGTCCAGCATCCCGT 1366
                          ||| :::|||      |||:::|||::::::   |||
Db        257 ----------------------Gly-AlaProGlyGlyLysLeuArgAlaAlaGlyPro- 268

Qy       1365 CTCCGTCGAAGTCGGTCACCACACCCCCTGTGCCCCGG---CCCTCAGGCTCCAAGGCG- 1310
                          :::   |||   |||:::   |||   |||:::|||:::   |||
Db        269 ----------ProLeuAlaThrGlyProLeuArgArgGlnProAlaGlyAlaThrAlaA 285

Qy       1309 ----------TCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGA 1261
                       ||||||      :::   :::|||      |||      ||||||
Db        285 rgLeuArgHisArgProGlyHisProAlaArgLeuArgArgGlnProAlaArgLeuArgA 305

Qy       1260 TGACGCGGAAGAGG---------------------------------------------- 1247
                :::   :::|||
Db        305 rgSerGlyGluArgProAlaGluSerProArgArgThrProGlyGluLeuArgGlyArgA 325

Qy       1246 ----------------CGGTTGGCTGAGGAGCTGCGGTAGG---------CAATGTTGT 1213
                           |||||||||   :::    ||||||        :::
Db        325 rgProAspAlaLeuProArgLeuAlaGlyLys-ProGlyArgTyrProAlaGluPheAla 344

Qy       1212 TGAAGAAGATCTCCAGCTCCTGGT--------------------------CATTGTCAA 1180
                    :::||||||||||||                          |||   :::
Db        345 ArgGlyAlaAlaProAlaProGlyProLeuAlaAlaAlaValProArgArgHisAlaArg 364

Qy       1179 AGT------------CGGCGGTGATGACCGTGCGGACAGGGGAGGGCATGGAGAACTTGG 1132
                :::          ||||||      |||   |||   ||||||||   ||||||
Db        365 AlaAlaGluProTrpArgArg------ProAlaGlyGlyGlyArgAlaAspArgThr--- 381

Qy       1131 GTGAGGCGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGCAGATAGAGGCGGT 1072
                          :::   |||:::   :::
Db        382 --------------AlaThrAlaProAlaArgProArgThrAlaCysAlaAlaAlaVal 396

Qy       1071 GGGGGCCATTCCAGTTGCCATAGACGATGTCCACTTTGCCATCACGGTTGAAGTCAGCCA 1012
                |||          |||   |||          |||   |||
Db        397 AlaGlyProArgHisAlaArgArgGluGlnProArgThrLeuProGlyProArgProPro 416

Qy       1011 GGGCGACACCTCGCCCATGCTGGTGGG---------------------------------- 985
                |||      |||   |||
Db        417 GlyAlaArgGlyThrHisLeuAlaGlySerAlaGlnProAlaAlaProAlaAlaAlaArg 436

Qy        984 ---------------------------------GGTCGTCCACACCAGCACTGGCCGCAG 958
                                              ||||||      |||      |||
Db        437 ProProAlaAlaLeuProAlaArgAlaArgArgGlyArgArgArgGlnPro---ProGly 455

Qy        957 CGTCCACAAAGGTGCCATCGCCCCGGTTGT----------------------------- 928
              ||||||  |||      |||||||||
Db        456 ArgProProArgArgLeuArgProGlyAlaProValAlaAlaAlaAlaProAlaThrGly 475

Qy        927 ------------------GGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATAT 886
                               |||   |||          |||      ||||||
Db        476 GlyAlaAlaThrGluProGlyAlaGlyAspSerArgHis---ArgCysArgSerGlyThr 494

Qy        885 CCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTATATTTGC 826
                |||
Db        495 AspArgPro---------------------------------------------------- 497

Qy        825 TGACCCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGAATGCCCCGGGAGAGGTCACTGG 766
                  :::|||:::||||||          |||          ||||||   |||   |||
Db        498 ---GlyArgProArgGlnProAla------AlaAlaArgAspProGlyProGlySerTrp 514

Qy        765 CCTCAGGGTCCATTTCAATGAGGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAA 706
                |||                          ||||||          ||||||
Db        515 Pro--------------------------GlyProSerAlaAlaArgArg--------- 522

Qy        705 TGTAGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAA 646
                                                                      :::
Db        523 ----------------------------------------------------AlaAlaArg 525

Qy        645 AGAGGCTGGCCACACCACGGGCCACGTTGACCTCATCGCTCAGGATGTCTTCCCACCGGT 586
              ||||||  ||||||
Db        526 ArgGlyAspProHisArgArgGlu---------------------------------- 533

Qy        585 TATTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTATTGG 526
                          |||      |||               |||||||||
Db        534 --------------ThrAlaLeuArgSerAlaAlaArgArgProArgArgArg------ 546
```

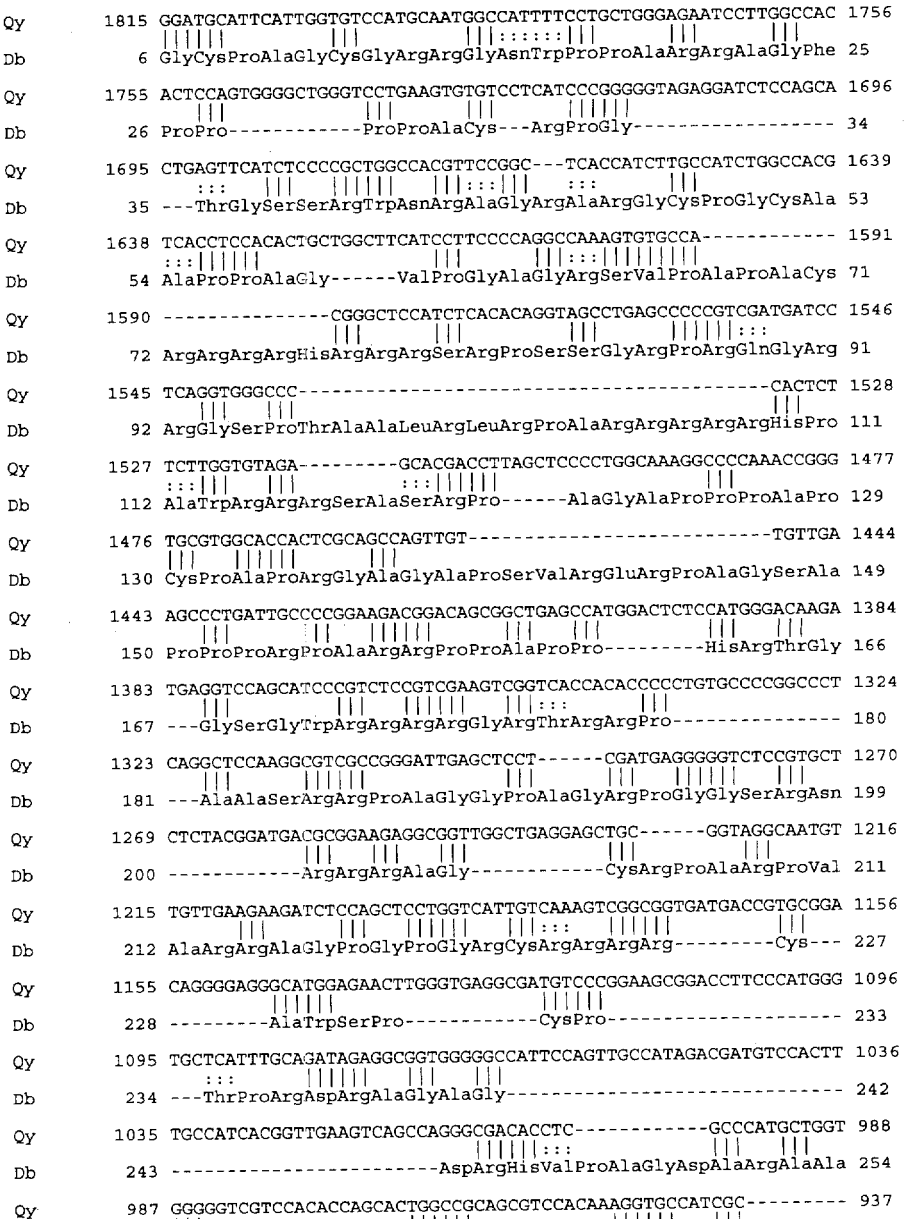


RESULT 3
US-09-252-991A-18296
; Sequence 18296, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18296
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18296

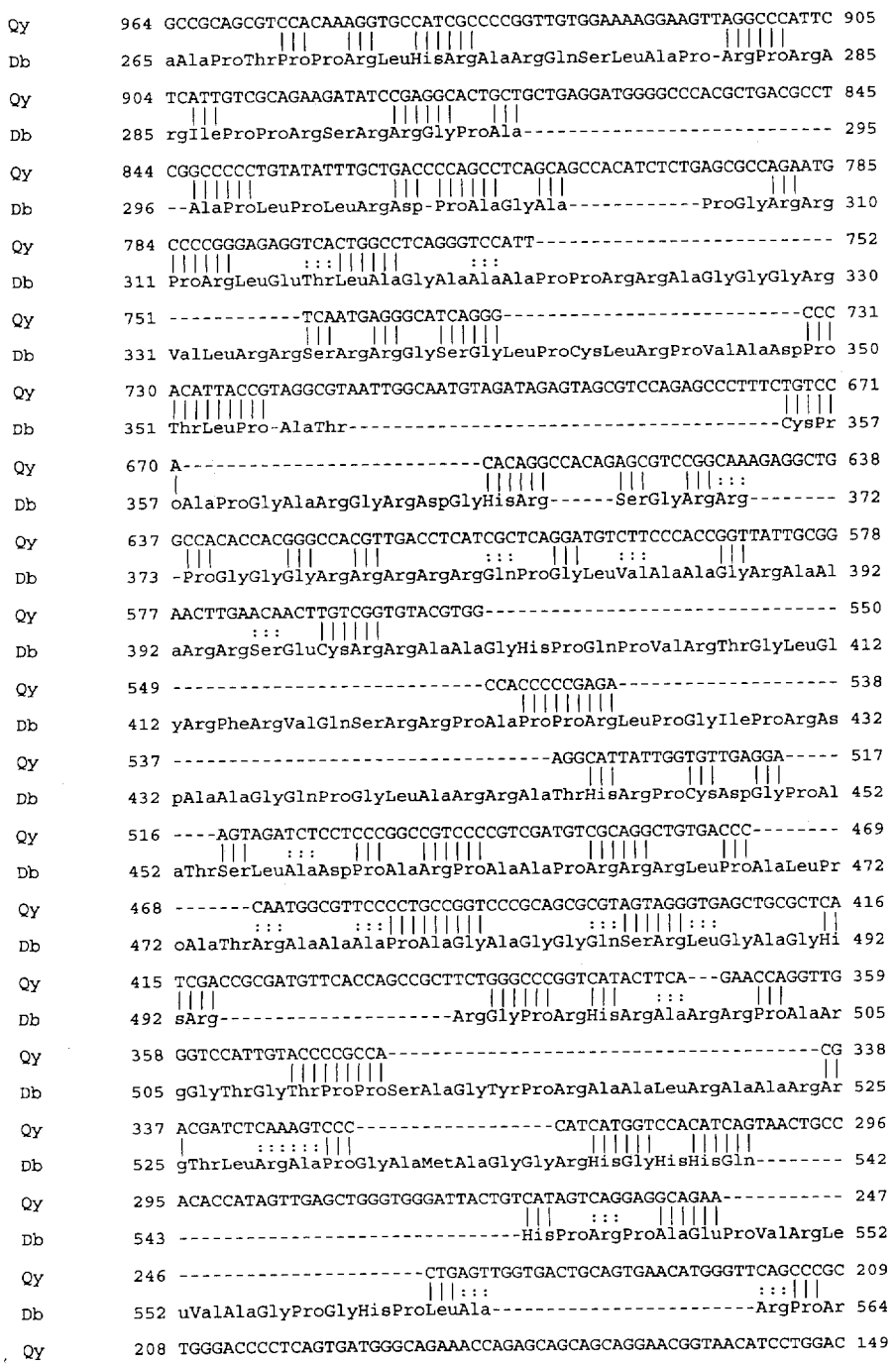
Alignment Scores:

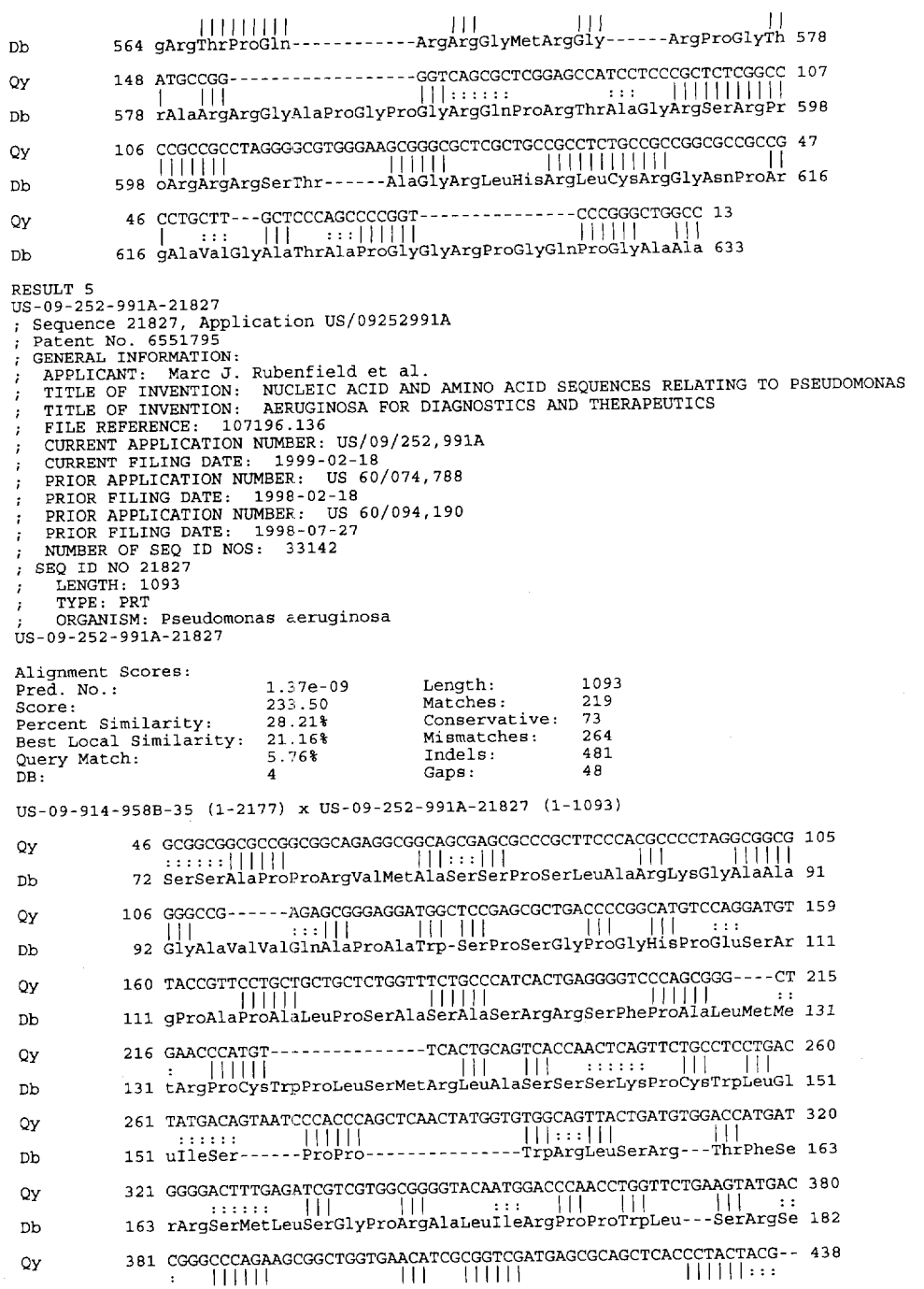
Pred. No.:	3.29e-10	Length:	467
Score:	239.50	Matches:	160
Percent Similarity:	32.81%	Conservative:	27
Best Local Similarity:	28.07%	Mismatches:	185
Query Match:	5.79%	Indels:	199
DB:	4	Gaps:	40

US-09-914-958B-35 (1-2177) x US-09-252-991A-18296 (1-467)



```
Db        255 Gly-----------------TrpPro---ThrGlySerArgCysAlaArgArgSerGly 267

Qy        936 ---CCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATATCCGAGG 880

Db        268 SerAlaGlyCysAla-------------------------------ArgProProArg 276

Qy        879 CACTGCTGCTGAGGATGGGGCCCACGCTG-----------------------ACGCCT 845

Db        277 ArgSerCysValArgTrpAlaAlaArgArgValAlaArgAspSerAlaArgArgSerAsp 296

Qy        844 CGGCCCCCTGTATATTTGCTGACCCCAGCCTCAGCAGCCACATCTCTG---AGCGCCAGA 788

Db        297 ArgPro-----------------ValArgSerAlaLeuArgGlyLeuProAlaAlaArg 310

Qy        787 ATGCCCCGGGAGAGGTCACTGGCCTCAGGGTCCATTTCAATGAGGGCATCAGGGCCCACA 728

Db        311 ArgSerArgProArgHisArgArgProGlySerValProProValArgThrSerArgAla 330

Qy        727 TTACCGTAGGCGT---------AATTGGCAATGT---------AGATAGAGTAGCGTCCA 686

Db        331 ThrPro-ArgArgTrpHisAlaArgTrpGlnAlaAlaIleArgArgProGlySerValLe 350

Qy        685 GAGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAAAGAGGCTGGCCACACCACGG 626

Db        350 uSerSerAlaThrProAlaIlePro---------GlyValArgGlyTyrProArgGluTy 367

Qy        625 GCCACGTTGACCTCATCGCTCAGGATGTCTTCCCACCGGTTATTGCGGAACTTGAACAAC 566

Db        367 rAlaArgAlaProArgProAla------LeuPro-------------------------- 376

Qy        565 TTGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTATTGGTGTTGAGGAAGTAGATCTCC 506

Db        377 -------------GlyArgProAlaAlaArg---------------------------- 382

Qy        505 TCCCGGCCGTCCCCGTCGATGTCGCAGGCTGTGACCCCAATGGCGTTCCCCTGCC----- 451

Db        383 -ProAlaArgProArgProCysArgArg------ProAlaAspArgSerProGlyGlnGl 400

Qy        450 ----------GGTCCCGCAGCGCGTAGTAGGGTGAGCTGCGCTCATCGACCGCGATGTTC 401

Db        400 yArgProProGlySerSerAlaThrArgArgTrpArgProCysHisArg---ArgAlaGl 419

Qy        400 ACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCATTGTACCCCGCC 341

Db        419 uProAlaAlaAlaProAlaGlyTyr---------ProGlyAlaIleAlaArgSerAlaTh 436

Qy        340 ACGACGATCTCAAAGTCCCCATCATGGT 313

Db        436 rArgArgThrHisAlaGlyArgArgGly 445

RESULT 4
US-09-252-991A-24567
; Sequence 24567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 24567
;   LENGTH: 681
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24567

Alignment Scores:

Pred. No.:                1.02e-09      Length:          681
Score:                    234.00        Matches:         194
Percent Similarity:       30.26%        Conservative:    36
Best Local Similarity:    25.53%        Mismatches:      242
Query Match:              5.66%         Indels:          288
DB:                       4             Gaps:            45

US-09-914-958B-35 (1-2177) x US-09-252-991A-24567 (1-681)

Qy       1801 GTGTCCATGCAATGGCCATTTTCCTGCTGGGAGAATCCTTGGCCA---CACTCCAGTGGG 1745

Db          1 ValThrHisAlaTrpPro------AsnTrpProProProTrpAlaCysAsnSerThrGly 18

Qy       1744 GCTGGGTCCTGAAGTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATC 1685

Db         19 ------------ArgThr-ProThrAspGlyArgSerGlySerThrArgArgSerSerAl 34

Qy       1684 TCCCCGCTGGCCACGTTCCGGCTCA---------------------------------- 1660

Db         34 aAlaCysTrpLysHisSerAlaIleArgArgArgValArgAsnArgSerAsnArgAlaTr 54

Qy       1659 ----------CCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCTTC 1610

Db         54 pProArgTrpProGlyCysAlaArgThrAlaProThrSerAspCysTrpSerAlaAsnAl 74

Qy       1609 CCCAGGC------------CAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGCCTGAG 1562

Db         74 aAlaSerArgTrpSerLysArgSerAlaProProAlaProArgAlaSerTrpSerThrAr 94

Qy       1561 CCCCCGTCGATGATCCTCAGGTGGGCCCCACTCTTCTTGG--------------TGTAG 1517

Db         94 gThrAlaArg------------GlySerProCysAlaTrpMetProLysAlaAlaCysAr 110

Qy       1516 AGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTCGCAGC 1457

Db        110 gArgArgPro-----------------------------------AlaAlaAlaMetAr 118

Qy       1456 CAGTTGTTGTTGAAGCCCTGAT---------TGCCCCGGAAGACGGACAGCGGCTGAGCC 1406

Db        118 gSerCysArgTrpSerSerGlyAsnGlyArgTrpArgTrpArgArgArgAlaValArgPr 138

Qy       1405 ATGGACTCTCCATGGGACAAGATGAGGTCCAGCATCC---CGTCTCCGTCGAAGTCGGTC 1349

Db        138 oTrp-----------------------ProAlaSerProArgValAlaArgValAlaGl 150

Qy       1348 A------CCACACCCCCTGTGCCCCGGCCCTCAGGCTCCAAGGCGTCGCCGGGATTGAGC 1295

Db        150 yValTrpProProSerLeuCysProAlaPro------ProArgArgArgArgProGlyAr 168

Qy       1294 TCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGATGACGCGGAAGAGGC----------- 1246

Db        168 gGlnArg-----------------------------ArgProArgGlyProAlaAlaGl 178

Qy       1245 ----------GGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGAAGATCTCCAGC 1196

Db        178 nArgArgAlaAlaTrpArgArgArgAlaGlyAspGln-----------------ProAl 192

Qy       1195 T------CCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAGGGG---AGGGC 1145

Db        192 aAlaArgProGlyArgGlyGlnArg-----------ThrCysLeuGlnProLeuLeuAl 208

Qy       1144 ATGGAGAACTTGGGTGAGGCGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGC 1085

Db        208 aValGlnProAlaValLeuGlnCysProAla---------ArgArgAlaGlyAspHisSe 225

Qy       1084 AGATAGAGGCGGTGGGGGCCATTCCAGTTGCCATAGACGATGTCCACTTTGCCATCACGG 1025

Db        225 rArgArgGlySerGlyGlyThrGlyAspProSerArgArgProGlyArgGlyAspGlyPr 245

Qy       1024 TTGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGTCGTCCACACCAGCACTG 965

Db        245 oProGlyGluProGlyAlaAspArgLeuAspSerGlyGlyGlyProAlaHisAlaProAl 265
```



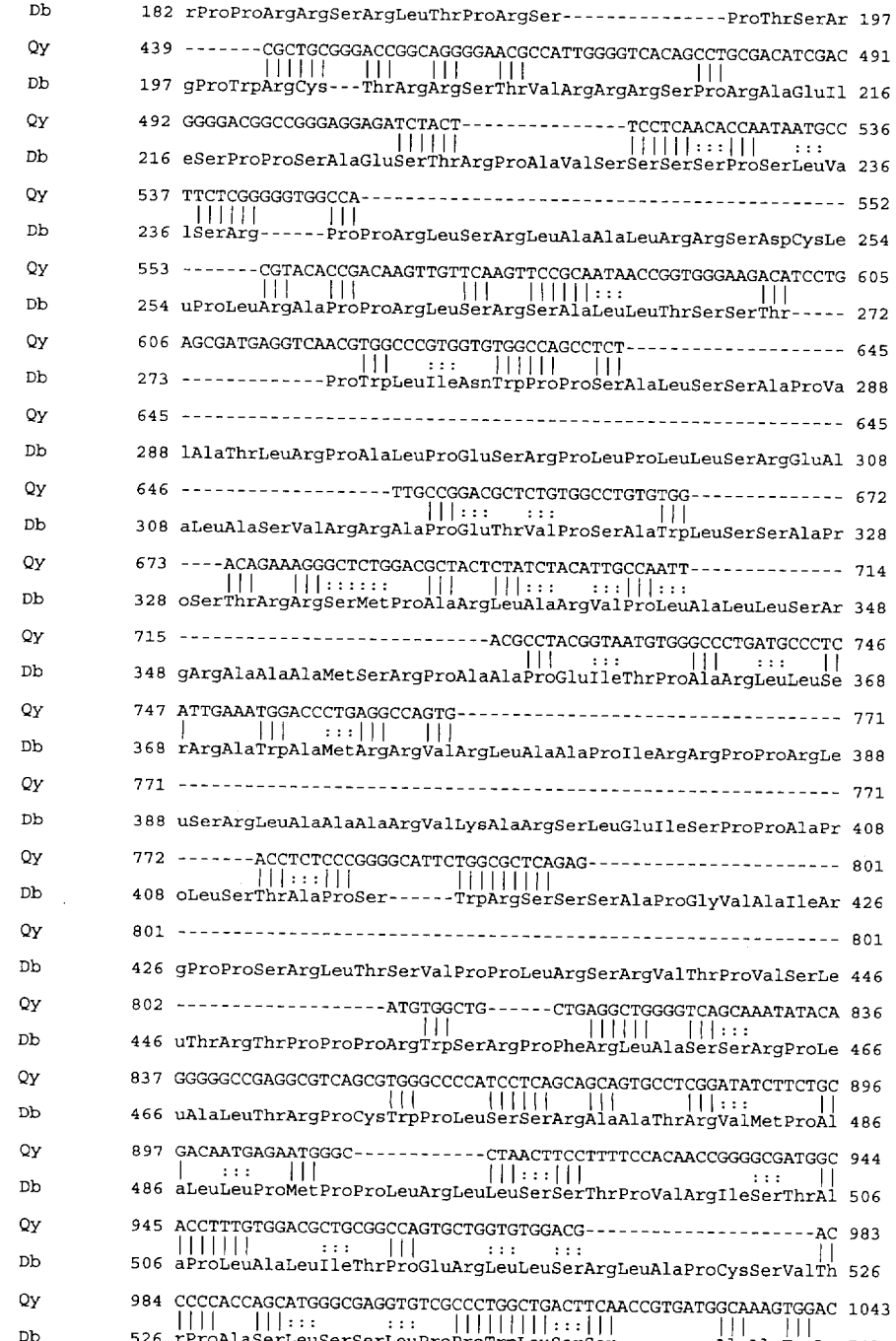


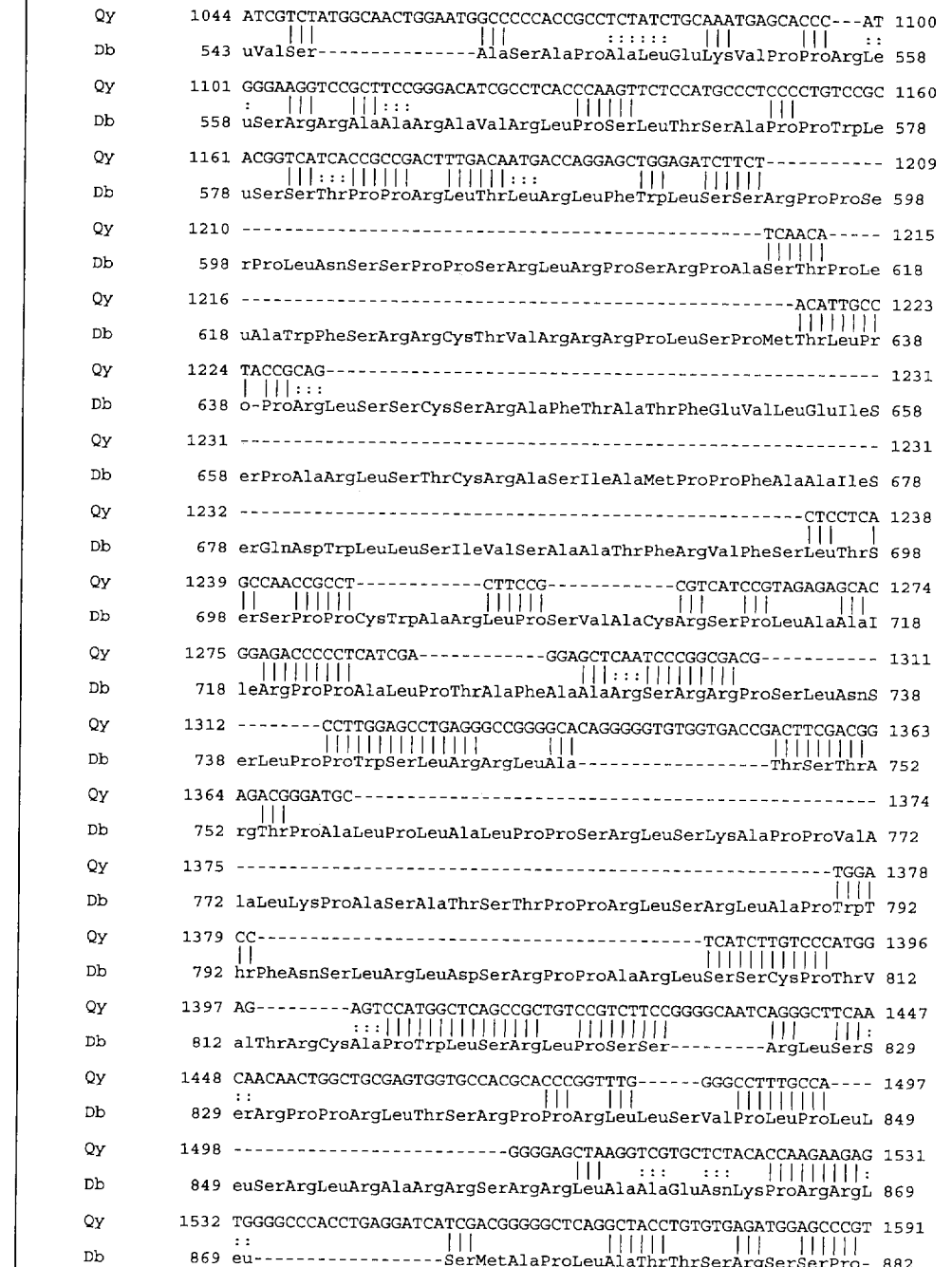
RESULT 5
US-09-252-991A-21827
; Sequence 21827, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21827
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21827

Alignment Scores:

Pred. No.:	1.37e-09	Length:	1093
Score:	233.50	Matches:	219
Percent Similarity:	28.21%	Conservative:	73
Best Local Similarity:	21.16%	Mismatches:	264
Query Match:	5.76%	Indels:	481
DB:	4	Gaps:	48

US-09-914-958B-35 (1-2177) x US-09-252-991A-21827 (1-1093)





```
Qy       1592 GGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGG 1651

Db        882 ------------------------------------------------------------ 882

Qy       1652 CAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTA 1711
                                                                        |||
Db        883 ------------------------------------------------------IleSerL 885

Qy       1712 CCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTC 1771
               |||          :::         |||:::            :::   :::   ||||
Db        885 euProProArgLeuSerArgLeuProThrArgAlaSerThrArgLeuProGluThrSerP 905

Qy       1772 CCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTG 1831
               |||||   ::::::     ||||||           |||   :::          |||
Db        905 roSerProArgLeuThrSerAlaTrp-----------AlaLeuThrArgValAspCysT 921

Qy       1832 CCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAA 1891
                 |||   :::   |||          |||:::   :::      |||      :::|
Db        921 rpLeuProSerArgPro------CysTrpProLeuSerSerValAlaAlaProThrSerA 939

Qy       1892 GAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCA---------------------- 1929
               |||||             |||   |||   ||||::::::
Db        939 rgSerPheArgProIleThrTrpProProArgLeuSerArgSerProArgValLeuThrT 959

Qy       1930 ----------------------CAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGAT 1966
                                     :::|||:::|||
Db        959 hrAlaArgSerArgLeuCysAsnArgProSerTrp------------------------- 970

Qy       1967 AGTGACACCACAAGTTGGGAAGAGCCTTG------------------------------- 1995
                        :::|||   |||   |||
Db        971 ----------ArgLeuSerArgLeuLeuAlaSerThrArgSerSerProSerAlaSerS 987

Qy       1996 --GTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCT 2034
                  :::|||      |||:::   |||         ||||||
Db        987 erLeuProProArgLeuSerArgCys---------ProPro 997

RESULT 6
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 30843
;   LENGTH: 663
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Alignment Scores:
Pred. No.:              2.04e-08        Length:         663
Score:                  217.00          Matches:        163
Percent Similarity:     29.69%          Conservative:   57
Best Local Similarity:  22.00%          Mismatches:     255
Query Match:            5.35%           Indels:         266
DB:                     4               Gaps:           37

US-09-914-958B-35 (1-2177) x US-09-252-991A-30843 (1-663)

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              |||   ||||||      |||             |||                |||||
Db         17 ArgAlaAlaArgAlaSerPro-----------TrpSerThrThrThrArgProProAla 32

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCA---------------------------- 90
                        :::   |||:::||||||   |||

Db         33 AspArgSerGlyAlaThrAlaArgValProAlaCysArgArgValProIleArgCysAla 52

Qy         91 --------------CGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCG 135
                           ||||||   ::::::      |||      ||||||
Db         53 AsnAlaTrpProGlyArgProMetSerSerThrSerArgArgProGlyTrp--------- 69

Qy        136 CTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCA 195
                 :::||||||   |||                              :::|||   |||
Db         70 ---SerProAlaValPro-----------------------AlaGlyAlaCys------ 78

Qy        196 CTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTC 255
                                                                    |||
Db         79 ------------------------------------------------------CysTrp 80

Qy        256 CTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACC 315
                 |||      |||         |||:::|||      |||
Db         81 LysThrArgAsnValAlaGlyArgSerAlaThrAspAlaTrpSer--------------- 95

Qy        316 ATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGT 375

Db         95 ------------------------------------------------------------ 95

Qy        376 ATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACT 435
                                                          ||||||      |||
Db         96 ------------------------------------------SerAlaProAlaProAla 101

Qy        436 ACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGG 495
                 |||                  |||   |||      ||||||:::
Db        102 AsnCysCysCysLeu---------ProProAlaGlyProSerProAlaSerProAlaPro 118

Qy        496 ACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGG--------- 546
                 |||   ||||||         :::            |||:::
Db        119 AlaAlaCysArgArgTrpProArgAlaAlaCysHisTrpProAlaSerAlaTrpTrpTrp 138

Qy        547 ---------------------TGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATA 585
                                    ||||||      |||      |||:::
Db        139 LeuAlaProAlaArgCysCysTrpProAlaProProAlaProAlaSerAlaGlyArgAla 158

Qy        586 ACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTG---------------- 630
                           :::         |||      ||||||
Db        159 CysCysAlaSerProAsnArgArgArgArgGluProTrpProProSerProTyrAlaSer 178

Qy        631 ---GTGTGGCCAGCCTCTTTG------CCGGACGCTCTGTGGCCTGTGTGGACAGAAAGG 681
                  |||||||||          |||   |||      ||||||
Db        179 ArgAlaGlyProAlaSerCysGlyArgProProAlaCysSerProVal------------ 194

Qy        682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATG 741
                  ||||||      |||          :::|||
Db        195 ---------AlaThrAlaProThrAlaThrCysSerPro--------------------- 204

Qy        742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
                 ||||||   :::         :::|||                  |||
Db        205 ProSerAlaArgSerAlaTrpLysProCysAlaCysAlaLysAla--------------- 219

Qy        802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCC 861
                          :::||||||||||           :::   |||:::||||||
Db        220 --------------ValGlySerAlaArg---------SerProAlaAsnAlaTrpPro 231

Qy        862 CCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCC 921
                 :::||||||      |||      :::||||||                    ::::::
Db        232 AlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgPro------------SerAla 247

Qy        922 TTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACG 981
                 ::::::|||         :::|||                  ||||||      |||
Db        248 IleAlaSerThrAlaArgArgSerProSerThrAsnGlyArgProAlaAlaCysProThr 267

Qy        982 ACC-----------------CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCA 1023
                 |||                 |||   |||                |||||
Db        268 ThrThrProProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProTrp--------- 284

Qy       1024 ACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATC 1083
                          |||      :::  :::      |||   |||:::   |||
```

```
Db        285 ---------------Trp---LysAlaProSerProAlaMetProProSerThrSerAla 298

Qy       1084 TGCAAATGAGCACCCATGGGAAGGTCC---------------GCTTCCGGGACATCGCCT 1128
Db        299 ThrArgLeuAlaAlaCysGlyHisAspGlyAspAlaGlyArgAlaSerGlnThrProTrp 318

Qy       1129 CACCCAAGTTCTCCATGCCCTCCC------CTGTCCGCACGGTCATCACCGCCGACTTTG 1182
Db        319 ArgAspThrSerProCysValProAsnCysValAsnTrpArgArgArgIlePro------ 336

Qy       1183 ACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT----------------- 1224
Db        337 ---------------TrpSerAlaValAlaLysThrSerProTrpProArgTrpProGly 351

Qy       1225 ---ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACC 1281
Db        352 MetProAlaGlyProArgProSerCysThrAlaValAlaAla---------------Trp 366

Qy       1282 CCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC---------------------- 1320
Db        367 AlaProAlaArgGlyGlyTyrAlaAlaAlaProArgAsnSerCysSerAlaGlyArgLeu 386

Qy       1321 ---------CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCG------------------ 1353
Db        387 ArgArgArgValArgArgSerAlaArgArgAlaTrpArgProTrpProAlaGlyArgAla 406

Qy       1354 ----------------------ACTTCGACGGAGACGGGATGCTGGACCTCATCTTGT 1389
Db        407 ThrProAlaSerAspAlaArgArgAspSerAlaSerArgGlyArgProAlaThrAlaAla 426

Qy       1390 CCCATGGAGAGTCCATGGCTCAGC------CGCTGTCCGTCTTCCGGGGCAATCAGGGCT 1443
Db        427 AspHisProAlaAlaTrpValSerAlaAlaArgArgThrSerSerAlaProIleAlaGly 446

Qy       1444 TCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAG 1503
Db        447 Ser--------------------------AlaProGlySerGlyThrAlaPro------ 455

Qy       1504 CTAAGGTCGTGCTCTACACCAAGAAGA------GTGGGGCCCACCTGAGGATCATCGACG 1557
Db        456 ------ArgCysHisProValArgLysAspGlyAlaGlyPro---------AlaSerThr 470

Qy       1558 GGGGCTCAGGCTACCTGTGTGAGATGG------AGCCCGTGGCACACTTTGGCCTGGGGA 1611
Db        471 GlyArgSerArgArg------ArgTrpAlaAlaAlaProAlaArgArgAlaArgThrGly 488

Qy       1612 AGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACG 1671
Db        489 IleArgArgThrAlaAla---------------------------------------- 494

Qy       1672 TGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACA--- 1728
Db        495 ------------------AlaAlaCysTyrProProProAlaProGlyArgArgThrPro 508

Qy       1729 ---------------CACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCC 1773
Db        509 AlaAlaArgArgArgSerAlaArgThrAlaProArgProAlaIleProAlaHisArgPro 528

Qy       1774 AGCAGGAAAATGGCCATTGCATGG-----------------ACACCAATGAATGCATCC 1815
Db        529 SerAlaSer---AlaThrGlyTrpProGlyProAlaLeuAlaCysProAlaAlaAlaGly 547

Qy       1816 AGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACA 1875
Db        548 ArgArgProSerProAlaProAlaProAlaProAlaProHisAlaArgProArgAlaPro 567

Qy       1876 GGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCT 1935
Db        568 ArgArgArgProGlyArgSerPro------AlaThrAlaProAlaAlaPheAlaAlaPro 585

Qy       1936 GCG 1938
Db        586 Ala 586
```

RESULT 7
US-09-252-991A-21881
; Sequence 21881, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21881
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21881

Alignment Scores:

Pred. No.:	2.76e-08	Length:	720
Score:	215.50	Matches:	191
Percent Similarity:	30.83%	Conservative:	39
Best Local Similarity:	25.60%	Mismatches:	261
Query Match:	5.31%	Indels:	255
DB:	4	Gaps:	41

US-09-914-958B-35 (1-2177) x US-09-252-991A-21881 (1-720)

```
Qy         20 CCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGCG 79
Db         92 ProGlyProArgSerGluArgArgArgGlnArgArgArgAlaTyrArgArgGlnGlyArg 111

Qy         80 CCCGCTTCCCACGCC---------------CCTAGGCGGCGGGGCCGAGAGCGGGAGGAT 124
Db        112 ProAlaGlyAspAlaTrpGlnProGlyProAlaArgTrpArgThrGlyGlnArgLysAsp 131

Qy        125 GGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTT 184
Db        132 ProGluGlyArg---GlyArgHis------AspArgGlnGlnProValLeuAla------ 146

Qy        185 TCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTC 244
Db        147 ---------------------------------AspArgArgGluTrpArgHisArgTyr 155

Qy        245 AGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTAC 304
Db        156 ArgGlyAlaAla---------------------------------------------- 159

Qy        305 TGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCT 364
Db        160 -----------------GlyArgGlnProArgArgArgAspPheGlnHisLeuGluGly 173

Qy        365 GGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGC------GGTCGATGA 418
Db        174 ArgArgGlu---------ArgProArgAlaThrGlyGlnProArgArgGlnGlyHisArg 190

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
Db        191 ArgGlnArgProAlaProHisArgAlaAlaProAla---GluProGlyGlnGlyGlyAla 209

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
Db        210 GlyArgAlaArgArgProGluProGlyArgArgArgThrValGlnArgArg--------- 226

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
Db        227 ---------------------ArgArgSerAlaArgGlnProGluGlnProGluArgGlu 239
```

```
Qy       599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
Db       240 ProArgArgArgAlaGlyGlnProGlyArgArgAlaGlyGlnArg------ArgGlnPro 257

Qy       659 TGTGGCCTGTGTGGACAGAAAGGGCTC---TGGACGCTACTCTATCTACATTGCCAATTA 715
Db       258 AspGlyAlaArgArgAlaProGlyGlnProTrpArgAspLeu--------------Leu 272

Qy       716 CGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCT 775
Db       273 GlnArgArgCysAlaGlyAlaAspGlnPro---------GlyArgAlaGlyGln------ 287

Qy       776 CTCCCGGGGCATTCTGGCGCTCAGAGATGT------------------------------ 805
Db       288 ---ProGlyArgGlnAlaAlaGlnArgCysArgArgAspAlaAlaGlyArgGlnProArg 306

Qy       806 ----------------------------GGCTGCTGAGGCTGG--------------- 820
Db       307 GlnGlnProPheArgArgAspGlnArgGlnGlyCysGlyGlyTyrProHisArgArgAla 326

Qy       821 GGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCT------------ 868
Db       327 GlyGlnGlnProGlnArgArgHisArgGlnGlnArgArgHisHisProGlyGlyArgPro 346

Qy       869 --------------------------CAGCAGCAGTGCCTCGGATATCTTCTGCGACAA 901
Db       347 AlaGlyGlnArgProAlaGlyThrGlyGlnArgGln------GlyProAlaArgArgGln 364

Qy       902 TGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGC 961
Db       365 ProGlu--------------ArgProArgProAlaArgArgArg---------ArgPro 376

Qy       962 GGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGG--------------GCGAGGTGT 1006
Db       377 GlyGlnArgAsnArgArgHisProArgProGlnTrpArgHisAlaGlyGlnProArgArg 396

Qy      1007 CGCCCT-----------------GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGT 1048
Db       397 ArgProAspArgHisAlaArgArgAlaAlaAlaAlaProAlaArgArgGlyGlyGlnArg 416

Qy      1049 CTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCA--------- 1099
Db       417 ArgArgArgGlyAsnLeuGlnArgProArgLeu----------HisProArgArgArg 432

Qy      1100 --------------------------------TGGGAAGGTCCGCTTCCGGGACATCGC 1126
Db       433 GlnProGlyGlnProArgArgAlaProAspArgArgArgGlnProAspProAlaHisArg 452

Qy      1127 CTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAA 1186
Db       453 ------------------------ProGlyProGly--------------------Gln 457

Qy      1187 TGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCG 1246
Db       458 GlnProGlyArgGlyAspLeuArgArgArgArg------ProGlyHisArgGlyArgSer 475

Qy      1247 CCT------------------------------------------------------- 1249
Db       476 ProGlyGlnGlnArgGlnGlyHisProGlyGlnProArgArgHisArgProAlaArgArg 495

Qy      1250 -----------CTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297
Db       496 ArgArgAlaGlyGlnProArgArgArgHisArgPheArgArgProPro------------ 511

Qy      1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357
Db       512 AspAlaArgGlnArgLeu-----------ProGlyGlnGlnArgGlnGlyProAlaLeu 527

Qy      1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTC------CATGGCTCA 1411
Db       528 ArgGlnArgArgProGluArgArgHis---------TrpArgAlaGlyGlnArgArgGly 544

Qy      1412 GCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCC 1471
Db       545 TrpProValAspGlnProGlyArgAlaGlyArgGlnGlnArgArgProArgGlnProTrp 564

Qy      1472 ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAG 1531
Db       565 ArgArgProGlnTrpGlnAlaValAlaAlaPro---GluArgArgGlnProGlyGlnPro 583

Qy      1532 TGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGT 1591
Db       584 TrpArgProAlaAspGlnArgArgArg----------------------------Thr 593

Qy      1592 GGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGG 1651
Db       594 GlyThr---------------------------------------------------- 595

Qy      1652 CAAGATGGTGAGCCGGAACGT---GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCT 1708
Db       596 ---AspGlyArgAlaArgArgLeuArgArgArgArgArgAsnLeuArgProGlyArgPro 614

Qy      1709 CTACCCCCGGGA------TGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCA 1762
Db       615 AlaProAspGlyArgAlaProGlyAlaThrProGlyProAlaGlyArgArgAlaArgArg 634

Qy      1763 AGGATTCTCCCAGCAGGAAAA-----------------TGGCCATTGCATGGACACCAA 1804
Db       635 GlnSerArgProAlaGlyArgArgProGlyGlnProGlyTrpPro---------AspGln 651

Qy      1805 TGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTA 1864
Db       652 CysSerArgProAlaGluHisArgAlaAlaGluArgProArg------------------ 665

Qy      1865 TGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAG---TCGGGGCTACGAGCCCAACGA 1921
Db       666 ---------GlnProProGlyArgArgAspPheGlnProAlaGlyLeuArgAlaAlaGly 682

Qy      1922 GGATGGCACAGCCTGCGT 1939
Db       683 GlnAlaHisArgGlnArg 688
```

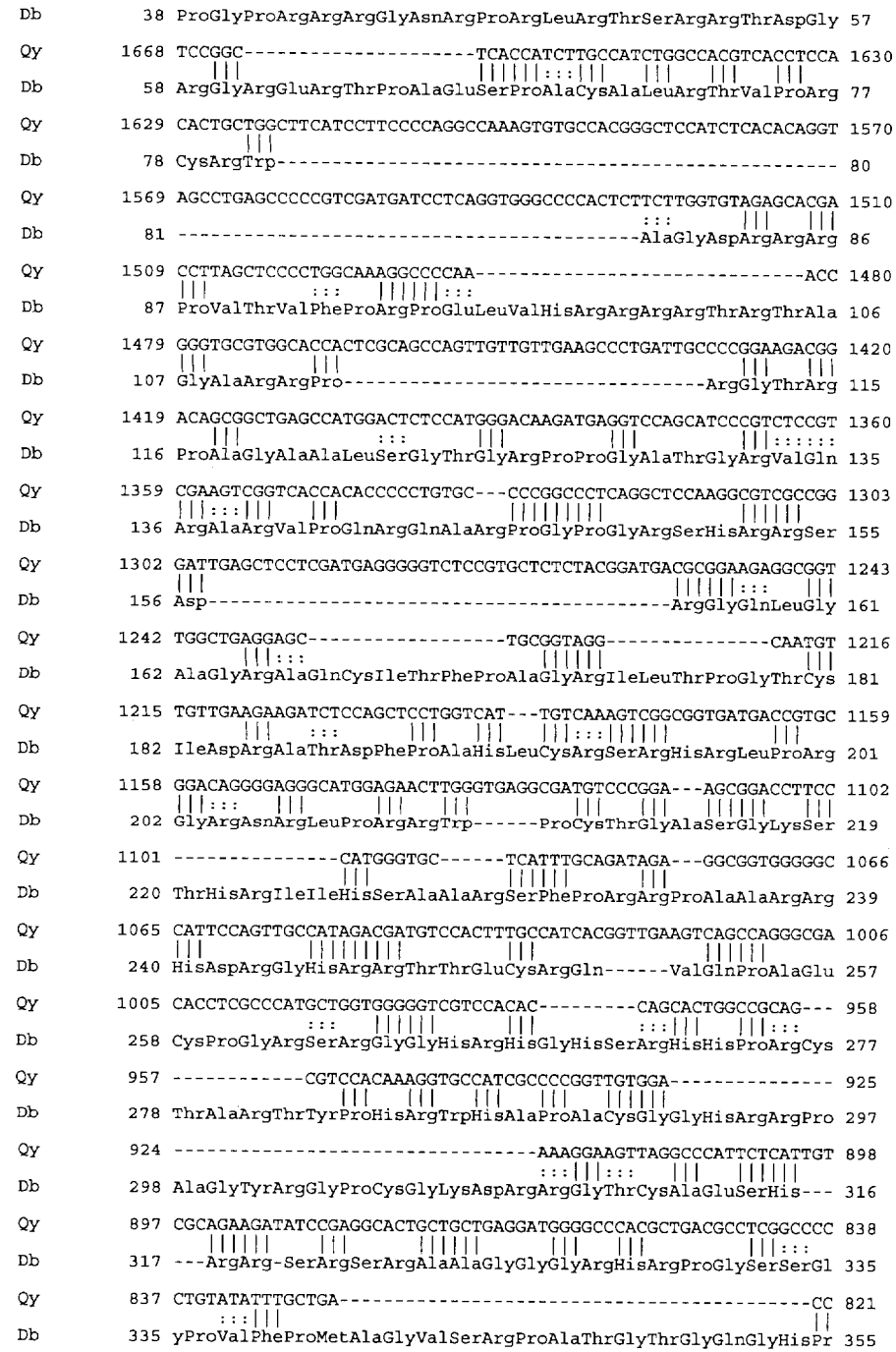
RESULT 8
US-09-252-991A-18995
; Sequence 18995, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18995
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18995

Alignment Scores:

Pred. No.:	2.77e-08	Length:	595
Score:	215.00	Matches:	165
Percent Similarity:	32.82%	Conservative:	48
Best Local Similarity:	25.42%	Mismatches:	223
Query Match:	5.20%	Indels:	214
DB:	4	Gaps:	37

US-09-914-958B-35 (1-2177) x US-09-252-991A-18995 (1-595)

```
Qy       1725 CCTCATCCCGGGGGTAGAGGATCT---CCAGCACTGAGTTCATCTCCCCGCTGGCCACGT 1669
```



RESULT 9
US-09-252-991A-19127
; Sequence 19127, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 19127
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19127

Alignment Scores:

Pred. No.:	4.3e-08	Length:	590
Score:	212.50	Matches:	163
Percent Similarity:	30.68%	Conservative:	26
Best Local Similarity:	26.46%	Mismatches:	216
Query Match:	5.14%	Indels:	211
DB:	4	Gaps:	36

US-09-914-958B-35 (1-2177) x US-09-252-991A-19127 (1-590)

```
Qy       1518 AGAGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTCGCA 1459
Db             7 ArgAlaArgPro--------------AspProAlaGlyGlyAlaAlaGlyGlyArgArg  21

Qy       1458 GCCAGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACT 1399
Db            22 AlaAsp---------ProProAlaGlyAlaGlyArgAlaAlaAlaThrLeuProAspPro  38

Qy       1398 CTCCATGGGACAAGATGAGGTCCAGCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCC 1339
Db            39 ------GlyAlaGluProAlaProGlyThrArgGlnArgArgGlyArgSerAlaGlnSer  56

Qy       1338 CTGTGC---CCCGGC------CCTCAGGCTCCAAGGCGTCGCCGGGATTGAGCTCCTCGA 1288
Db            57 GlnAlaArgProGlyGlyArgProAlaAlaProGlyArgThrArgAla---AlaArgArg  75

Qy       1287 TGAGGGGGTCTCCGTGCTCTCTACGGATGACGCGGAAGAGGCGGTTGGCTGAGGAGCTGC 1228
Db            76 ProArgArgLeuAlaAlaProGlyGlyGlyAlaHisArgGlyAlaTrpArgAspGlnPro  95

Qy       1227 GGTAGGCAATGTTGTTGAAGAAGATCTCCAGCTCCTGGTCATTGTCAAAGTCGGCGGTGA 1168
Db            96 GlyAlaGln---------ProGluArgGlyAlaProGlyHisGly--------------- 107

Qy       1167 TGACCGTGCGGACAGGGGAGGGCATGGAGAACTTGGGTGAGGCGATGTCCCGGAAGCGGA 1108
Db           108 ---ProGlyGlyGln------------ArgThrGlyVal---------ValGlyLeuGly 119

Qy       1107 CCTTCCCAT------GGGTGCTCATTTGCAGATAGAGGCGGT------------------ 1072
Db           120 ProGlyGluArgProGlyProAlaPheLeuSerArgGlyGlyLeuGlnArgArgArgGly 139

Qy       1071 ------GGGGGCCATTCCAGTTGCCATAGA------------CGATGTCCACTTTGCCAT 1030
Db           140 ArgSerArgGlyLeuSerGlnAlaAspArgLysHisProProArgArgProAlaAlaHis 159

Qy       1029 CACGGT---------------------------------------------------- 1024
Db           160 ProSerArgProArgArgThrSerGlnGlyAlaGlnArgAlaValProGlyGlyValPro 179

Qy       1023 TGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGTCGTCCA---CACCAGCAC 967
Db           180 ArgAlaArgProGlyGlyArgLeuAlaLeuAlaGlyGlySerTrpProGlyHisGlySer 199

Qy        966 TGGCCG---------------------------------CAGCGTCCACAAAGGTGCCAT 940
Db           200 ArgProAlaGlyThrGlyAspAlaAspAlaArgHisAlaGlnArgHisGlnArgAlaGln 219

Qy        939 CGCCCCGGT------------TGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGA 892
Db           220 AlaProGlyArgAlaAlaAlaThrGlyArgAspSerValArgGlyGlyGlnArgGluHis 239

Qy        891 AGATATCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTAT 832
Db           240 ArgAspProArgSerGlyLeuGlnProAlaGlyGlyGln-------------------- 252

Qy        831 ATTTGCTGACCCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGAATGCCCCGGGAGA--- 775
Db           253 ---------ProGlyLeuGlnSerHisHisArgLeuSerProArgGlyArgGlyArgAla 269

Qy        774 --------------GGTCACTGGCCTCAGGGTCCATTTCAA------------------ 748
Db           270 GlnArgGlnProAlaGlyGluTyrProArgGlyProAlaProLeuProAlaAspProArg 289

Qy        747 ------------------------TGAGGGCATCAGGGCCCACATTACCGTAGGCGTAAT 712
Db           290 ArgAlaAlaProAlaArgAlaValAlaGlyArgThrGlyGlyTyr------AlaGluGlu 307

Qy        711 TGGCAATGTAGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAGCGTC 652
Db           308 TrpArgSer--------------LeuSerProMetAlaAlaAlaGlnProAlaAlaArg 322

Qy        651 CGGCAA---AGAGGCTGGCCACACCACGGGCCACGTTGACCTCATCGCTCAGGATGTCTT 595
Db           323 ArgGlnTrpArgGlyAspProHisArg------ArgPheLeuHisArgSerGlnCys--- 339

Qy        594 CCCACCGGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGG 535
Db           340 ------------------------------------------------ProProArgArg 343

Qy        534 CATTATTGGTGTTGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGTCGATGTCGCAGGCTG 475
Db           344 --------------------ArgGlyAlaProAlaSerProAlaAlaLeuArgArgThr 356

Qy        474 TGACCCCAATGGCGTTCCCCTGCCGGTCCCGCAGCGCGTAGTAGGGTGAGCTGCGCTCAT 415
Db           357 AspArgProGlyGlnProProValValProArgAlaAlaAlaArgGlyGlyProAlaGly 376

Qy        414 CGACCG---------CGATGTTCACCAGCCGCTTCTGGGCCCGGTCATACTTCA------ 370
Db           377 AlaProGlyGlyGlyArgAlaGlyProAlaAlaAspArgProGlyProLeuGlnAlaAla 396

Qy        369 ------GAACCAGGTTGGGTCCATTGTACCCCGCCACGACGATCTCAAAGTCCCCATCAT 316
Db           397 GlnArgGlnProArgProArgGlyArgArgProValAlaGlnGlyAspGlyProAlaPro 416

Qy        315 GGTCCACATCAGTAACTGCCACACCATAGTTGAGCTGGG---------------------- 277
Db           417 AlaAlaHisArgAlaGlySerGlnHisProGlyAlaSerGlyArgArgArgValArgHis 436

Qy        276 ---------TGGGATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTGA 226
Db           437 SerProArgTrpArgHisArgHisArgGlyPro--------------------------- 447

Qy        225 ACATGGGTTCAGCCCGCTGGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGCAGGA 166
Db           448 ------ValAlaProGlyArgThrPro--------------------------AlaGly 456

Qy        165 ACGGTAACATCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCGGCCC 106
Db           457 ProValAlaPro---------AlaGlyGluArgAlaArgPro------ArgValAspPro 471

Qy        105 CGCCGCCTAGGG------------GCGTGGGAAGCGGGC---------------------  79
Db           472 ArgArgLeuAlaGlyAspGlnProValPheArgAlaGlyThrGlyAsnLeuArgAlaAsp 491

Qy         78 --------------------GCTCGCTGCCGCCTCTGCCGCCGGCGCCGCCGCCTGCTT  40
Db           492 GluProGlyGluProGlyAspAlaThrCysGlnAlaThrArgArgGlnTyrLeuProVal 511

Qy         39 GCTCCC--------------------------AGCCCCGGTCCCGGG  19
Db           512 LeuProArgGlnProAlaGlyGlnHisProArgThrProAlaProGly 527
```

RESULT 10
US-09-252-991A-27068
; Sequence 27068, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27068
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27068

Alignment Scores:

Pred. No.:	4.45e-08	Length:	638
Score:	212.50	Matches:	188
Percent Similarity:	29.81%	Conservative:	26
Best Local Similarity:	26.18%	Mismatches:	263
Query Match:	5.14%	Indels:	241
DB:	4	Gaps:	42

US-09-914-958B-35 (1-2177) x US-09-252-991A-27068 (1-638)

```
Qy       1830 ACACGAATGGGAACTGGATGCATTCATTGGTGTCCATGCAATGGCCATTTTCCTGCTGGG 1771
Db         35 ThrArgHisGlyAlaAlaLeuProValAlaArgProValProGlyAlaValAlaAlaLeu 54

Qy       1770 AGAATCCTTGGC---------CACACTCCAGTGGGGCTGGGTCCTGAAGTGTGTCCTCAT 1720
Db         55 ArgArgThrGlyArgAlaGlnHisProProVal--------------------ProArg 67

Qy       1719 CCCGGGGGTAGAGGATCTCCAGCA---CTGAGTTCATCTCCCCGCTGGCCACGTTCCGGC 1663
Db         68 GlnArgArgArgArgAlaProAlaGlyArgAlaAlaSerProGlyLysProArgArgGly 87

Qy       1662 TCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCTTCCCCAGGC 1603
Db         88 LeuAlaGly---------AlaThrAlaGlyPro------------ArgProAspPro--- 99

Qy       1602 CAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGCCTGAGCCCCCGTCGATGATCCTCA 1543
Db        100 ArgAlaLysProArgGlyValAlaArgIleGlnArgArgArgProAlaArgArgArgGln 119

Qy       1542 ---------------GGTGGGCCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTCCCC 1498
Db        120 AlaArgArgArgProAlaGlyProAlaAlaThrGly---ArgAlaArgPro--------- 135

Qy       1497 TGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGAAGCCCT 1438
Db        136 ------HisArgGlnProGlyThrMetValArgLeu------------------------ 145

Qy       1437 GATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGATGAGGT 1378
Db        146 ------ProGlyArgArgGlyGlyAlaAlaPro------------GlyArgArg---Leu 158

Qy       1377 CCAGCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCCCTGTGCCCCGGCCCTCAG--- 1321
Db        159 ProHisProAlaAlaAlaArgGlnProAlaProAspProProGlnProGlyProArgGln 178

Qy       1320 ----------------------GCTCCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGG 1282
Db        179 AlaGlyGlyLysArgGlnProGlyProGlyArgArgArg---------------------- 191

Qy       1281 GGTCTCCGTGCTCTCTACGGATGACGCGGAAGAGGCGGT------------TGGCTGAGG 1234
Db        192 ------ArgAlaArgIleGlyGlnProAlaArgAlaGlyArgAlaAspArgGlyLeuAla 209

Qy       1233 AGCTGCGGTAGGCAATGTTGTTGAAGAAGATCTCCAGCTCCTGGTCATTGTCAAAGTCGG 1174
Db        210 AlaAlaArgArg------------AlaArgArgAlaThrLeuGlyGlnArgArgLeuArg 225

Qy       1173 CGGTGATGACCGTGCGGACAGGGGAGGGCATGGAGAACTTGGGTGAGGCGATGTCCCGGA 1114
Db        226 Arg------HisAspGlyAlaGlyArgArgArgArgGlyTrpArgArgGlnCys------ 241

Qy       1113 AGCGGACCTTCCCATGGGTGCTCATTT-------------------------------- 1087
Db        242 ArgArgProArgArgGlyAlaAlaAlaArgThrArgGlnProAlaAlaThrLeuSerGly 261

Qy       1086 --------------------GCAGATAGAGGCGGTGGGGGCCATTCCAGTTGCCATAGA 1048
Db        262 ArgValAlaProHisProArgProAspArgAlaSerProAlaThrValArgArgHisArg 281

Qy       1047 CGA--------------TGTCCACTTTGCCATCACGGTTGAAGTCAGCCAGGG------ 1009
Db        282 ArgGlnThrGlyArgGlyAlaPro-----------GlyProGlyHisProGlyThrAla 297

Qy       1008 ---------CGACACCTCGCCCATGCTGGTGGGGGTCGTCCACACCAGCACTGGCCGCAG 958
Db        298 GlyAlaArgArgThrProAlaProAlaArgProGlyThrProAspProGlyTrpAspAsp 317

Qy        957 CGTCCACAAAGGTGCCATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGT 898
Db        318 ArgProAspProAlaHisArgPro---------------------------------- 325

Qy        897 CGCAGAAGATATCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGC--------- 847
Db        326 ArgHisArgGlnProAlaLysThrPro---GlyAlaGlyProArgAlaThrGlyAlaGly 344

Qy        846 ---------CTCGGCCCCCTGTATATTTGCTGACCCCAG---CCTCAGCAGCCACATCTC 799
Db        345 AlaLeuGlyLeuGlyArgArgArgLeuLeuAlaProHisIleProAlaTyrProHisArg 364

Qy        798 TGAGCGCCAGAATGCCCCGGGAGAGGTCACTGGCCT------CAGGGTCCATTTCAATGA 745
Db        365 ---GlyProAlaGlnProGlyArgPheAlaGluProProAlaGlnLeuProArgArgAla 383

Qy        744 GGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAATGTAGATAGAGTAGCGTCCAG 685
Db        384 GlyArgArgAspPro---------------------------------------------- 388

Qy        684 AGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAAAGAGGCTGGCCACACCACGGG 625
Db        389 ------------ProProArgArgThrGlyArgArgGlnGlnProAspProGlyArgGly 404

Qy        624 CCACGTTGACCTCATCGCTCAGGATGTCTTCCCACCGGTTATTGCGGAACTTGAACAACT 565
Db        405 GlnGlnArgProAlaArgArgGlyArgAlaProGly------------------ArgArg 418

Qy        564 TGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTATTGGTGTTGAGGAAGTAGATCTCCT 505
Db        419 TyrArgAlaAspProArgArgProArgArgHis-----------GlySerGlyAspPro 434

Qy        504 CCCGGCCGTCCCCGTCGA------TGTCGCAGGCTGTGACCCCAA--------------- 466
Db        435 AlaGlyArgArgArgArgGlnProAspArgArgArgGlnSerGlnArgArgProGlySer 454

Qy        465 ------TGG---------------CGTTCCCCTGCCGGTCCCGCAGCGCGTAGTAGGGTG 427
Db        455 GlyAlaTrpProAlaGlyAspArgArgGluProGlyArgProAlaArgAlaGlyArgArg 474

Qy        426 AGCTGCGCTCATCGACCGCGATGTTCACCAGCCGCTTCTGGGCCCGGT------------ 379
Db        475 GlyAlaArgGlnArgProValAspArgAlaProGlyArgGlyValGlyAspHisArgGln 494

Qy        378 --------------------------CATACTTCAGAACCAGGTTGGGTCCATTGTACC 346
Db        495 ArgAlaAspGlyAspProLeuAspArgArgThrAspGlnProAlaArgProGlnCysArg 514

Qy        345 ---------CCGCCACGACGATCTCAA----------------------------AGT 325
Db        515 HisArgGlyArgProArgArgArgProGlyProArgLeuArgArgGlyArgArgGlySer 534
```

```
Qy        324 CCCCATCATGGTCCACATCAGTAACTGCCACACCATAGTTGAGCTGGGTGGGATTACTGT 265
Db        535 ProLeuAlaGlyProAlaHisArgArgArgHisArg---------GlyAspProAlaAla 551

Qy        264 CATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTGAACATGGGTTCAGCCCGCTGGG 205
Db        552 His-------------------------------------------------ArgProProAla 556

Qy        204 ACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGCAGGAACGGTAACATCCTGGACATGC 145
Db        557 Thr------------------GlyArgAlaAlaIleGly---------------------- 563

Qy        144 CGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCGGCCCCGCCGCCTAGGGGCGTGGGAA 85
Db        564 ---GlyGlyHisAlaGlnProGlyArgAlaCysArgThrHisArgArgThrGlyArgArg 582

Qy         84 GCGGGCGCTCGCTGCCGCCTCTGCCGCCGGCGC-------------------------- 52
Db        583 GlyGlyArgArgAlaGlyArgSerGlyArgArgAspProTyrHisArgArgAspGlyArg 602

Qy         51 ------CGCCGCCTGCTTGCTCCCAGCCCCGGTCCCGGGCTGGCCTCGAGCCTC 4
Db        603 ThrHisArgArgArgLeuAspProAlaGluProGlyGlyArgArgAspSerLeu 620
```

RESULT 11
US-09-252-991A-17231
; Sequence 17231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17231
; LENGTH: 2294
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17231

Alignment Scores:

Pred. No.:	8.41e-08	Length:	2294
Score:	212.00	Matches:	193
Percent Similarity:	30.87%	Conservative:	33
Best Local Similarity:	26.37%	Mismatches:	273
Query Match:	5.13%	Indels:	234
DB:	4	Gaps:	36

US-09-914-958B-35 (1-2177) x US-09-252-991A-17231 (1-2294)

```
Qy       1779 CCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTGGGT---CCTGAAGTGTGTCCT 1723
Db          1 ProArgGlyValValArgArgHisValProAlaAlaThrAlaPheProAlaAlaArgPro 20

Qy       1722 CATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGGCCACGTTCCGGC 1663
Db         21 LeuProGlyAsnProGlyGly-----------------------TrpProLeuProLeu 32

Qy       1662 TCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCTTCC---CCA 1606
Db         33 AlaProGlyGlyArgLeuGlyLeuPheArgAlaGluAlaGlyArgCysProAlaAspPro 52

Qy       1605 GGCCAAAGT---------------GTGCCACGGGCTCCATCTCACACAGGTAGCCTGAGC 1561
Db         53 GlyGlnAlaGlyHisArgGlnArgProArgArgAlaAlaGlyTrpTyrAlaValGlnHis 72

Qy       1560 CCCCGTCGATGATCCTCAGGTGGGCCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTC 1501
Db         73 ProArgArgArgSerGlyAlaGly------AlaAlaGluCysArgGlyLysPro---Ala 89

Qy       1500 CCCTGGCAAAGG---CCCCAAACCGGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGA 1444
Db         90 ArgTrpHisArgLeuProArgProAlaAlaGluAlaGlyArgHisProArgHisGlyArg 109

Qy       1443 AGCCCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGA 1384
Db        110 GlyGlyArgGlyProArgArgArgProLeuArgGlnProGlyAla--------------- 124

Qy       1383 TGAGGTCCAGCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCCCTGTGCCCCGGC--- 1327
Db        125 ---GlySerAlaArgArgGlnCysArgLeuArgGluProArgProGlyGlyProGlyGln 143

Qy       1326 ---------CCTCAGGCTCCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTC 1276
Db        144 ProGlyValProProAlaAspArgAlaArgArgProArgArgIleArgArgGlyGlyArg 163

Qy       1275 CGTGCTCTCTACCGATGACGC---------------------------GGAAGAGGCGGT 1243
Db        164 ArgAla------GlyProArgGluAlaArgGlyProGlyProGluProGlyArgGly--- 180

Qy       1242 TGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGAAGATCT----------------- 1201
Db        181 ---LeuProAlaAlaGlyAlaArgHisCysArgArgAlaAlaProGlyValLeuArgGln 199

Qy       1200 -----------CCAGCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAG 1153
Db        200 GlyAlaAlaGlyProAlaProProAlaArgArgSerProArgArgAlaGlyGlyGlyIle 219

Qy       1152 GGGAGGGCATGGAGAACTTGGGTGAGGCGATGTCCCGGAAGCGGA---------CCTTCC 1102
Db        220 GlnArgArgGlyArgLeuGlyAsnArgArgLeuProAlaAlaAlaThrGlyGlnProArg 239

Qy       1101 CATGGGTGCTCATTTGCAGAT---------------------------------------- 1081
Db        240 ArgGlyProHisTrpSerProValProArgAlaProAlaAlaProAlaGlyAlaValPro 259

Qy       1080 ---AGAGGCGGTGGGGGCCATTCCAGTTGCCAT---AGACGATGTCCACTTTGCCATCAC 1027
Db        260 ArgThrGlyGlyProGlyArgArgValLeuTyrLeuHisArgGlnProAlaAlaSerArg 279

Qy       1026 GGT---------------TGAAGTCAGCCAGGGCGACACCTC------------------ 1000
Db        280 GlyValProGlyValGlnProArgArgProGlyArgGlnLeuArgProGlyAspAlaArg 299

Qy        999 -----------GCCCATGCTGGTGGGGGTCGTCCACACCAGCACTGGCCG--------- 961
Db        300 ProAlaProGlyAlaGlyAspGlyAspArgArgGlnGlyAspArgHisProAspArgArg 319

Qy        960 ----------------------------CAGCGTCCA-----------CAAAGGTGC 943
Db        320 GluGlyLeuArgArgAlaArgArgAspAlaGlnArgProGlnProAspArgGlnArgGln 339

Qy        942 CATCGCCCCGGT---------------------------------------------- 931
Db        340 LeuArgProGlyGlyAlaArgArgArgProArgArgGluProAlaArgAlaLeuProAla 359

Qy        930 TGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATATCCGAGGCACTGCTGC 871
Db        360 AlaGlyGlnGlyArgLeuGlyGlyAlaHisProArg---------Pro-GlyGlnProAl 376

Qy        870 TGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTATATTTGCTGACCCCAGCCTCAG 811
Db        376 aThrAlaGlyTyrArgProAlaAlaAspAlaGlyGlyAlaAlaAlaAsp-ProLeuProA 396

Qy        810 CAGCCACATCTCTGAGCGCCAGAATGCCCCGGGAGAGGTCACTGGCCTCAGGGTCCATTT 751
Db        396 rgArgLeuHisLeuPheArgThrGlyProArgGlnArgAlaLeuGluAlaAlaAspAlaL 416

Qy        750 CAATGAGGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAATGTAGATAGAGTAGC 691
```

```
Db        416 euArgArgLeuArgLeuPro---------------------------------------- 422

Qy        690 GTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAAAGAGGCTGGCCACAC 631
                                                                  |||      |
Db        423 ----------------------------------------------HisArgArgProValP 428

Qy        630 CACGGGCCACG------TTGACCTCATCGCTCAGGATGTCT---------TCCCAC---C 589
              ||||||||         |||          |||:::|||         ||||||   |
Db        428 roArgAlaGluProGlyLeuLeuGlyHisSerArgIleSerAsnAspAlaSerHisArgA 448

Qy        588 GGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTAT 529
              ||   ||||||:::              |||          |||      |||   |||:::
Db        448 rgSerLeuArgHisProGlyArgArg-ArgGlnArgLysProAlaArgArgProHisHis 467

Qy        528 TGGTGTTGAGGAAGTAGATCTCCT---CCCGGCCGTCCCCGTCGATGTCGCAGGCTGTGA 472
                    ||||||   |||   ||||||      ||||||
Db        468 AspHisAlaAlaSerArgArgProArgProGlyThrGlyArgArg--------------- 482

Qy        471 CCCCAATGGCGTTCCCCTGCCGGTCCCGCAGCGCGTAGTAGGGTGAGCTGCGCTCATCGA 412
              |||      |||              ||||||      :::          |||   |||
Db        483 ProProArgArgArgSerLeuHisAlaAlaAlaAspGlyGlnProGlyAlaAlaGluArg 502

Qy        411 CCGCGATGTTCACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCAT 352
              ||||||      |||      :::|||   |||              |||      |||
Db        503 ProArgArgArgProGlnProAlaGlyAlaGlyArgArgProThrAlaGlyAspAlaHis 522

Qy        351 TGTACCCCGCCACGACGATCTCAAAGTCCCCATCATGGTCCACATCAGTAACTGCCACAC 292
                    ||||||            :::   ||||||   |||
Db        523 ------ProPro------------AlaGlnHisHisArgPro------------------ 530

Qy        291 CATAGTTGAGCTGGGTGGGATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTG 232
                  |||   ||||||
Db        531 ---SerGlyAlaGly------------------------------------------- 534

Qy        231 CAGTGAACATGGGTTCAGCCCGCT---------------------------------- 208
                           |||||||||
Db        535 ---------------GlnProAlaArgAlaThrAlaArgLeuProAlaProValArgGly 549

Qy        207 ---GGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGCAGGAACGGTAAC-ATCCTG 152
                 |||||||||                  |||:::   ||||||   :::      |||
Db        550 GluGlyThrProGlyArgArgGlyAlaGlyArgSerAspAlaGlyAlaLeuArgAlaLeu 569

Qy        151 GACATGCCGGGGTCAGCG--------------------------------CTCGGAGCC 125
                    |||      :::                               ||||||
Db        570 HisArgProArgArgSerGlyAlaGlnHisAlaValGlyGlnHisGlnArgLeuGlyGln 589

Qy        124 ATCCTCCCGCTCTCGGCCCCGCCGCCTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCG--- 68
                 |||          |||      ||||||:::|||             ||||||
Db        590 AlaGluProAlaAspHisProAlaGlnArgGlyLeuGlyArgArgGluGlyLeuProVal 609

Qy         67 ------------CCTCTGCCGCCGGCGCCGCCGCCTGCT-------------------- 41
                        |||      |||||||||   ||||||
Db        610 AlaArgThrLeuProAlaGluProAlaProAlaProAlaProAlaGlyAlaAlaValPro 629

Qy         40 TGCTCCCAGCCCCGGTCCCGGGCTGGCCTC 11
                 |||||||||   |||      |||
Db        630 ValHisGlnProArgLeuArgArgProLeu 639
```

RESULT 12
US-09-252-991A-23774
; Sequence 23774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23774
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23774

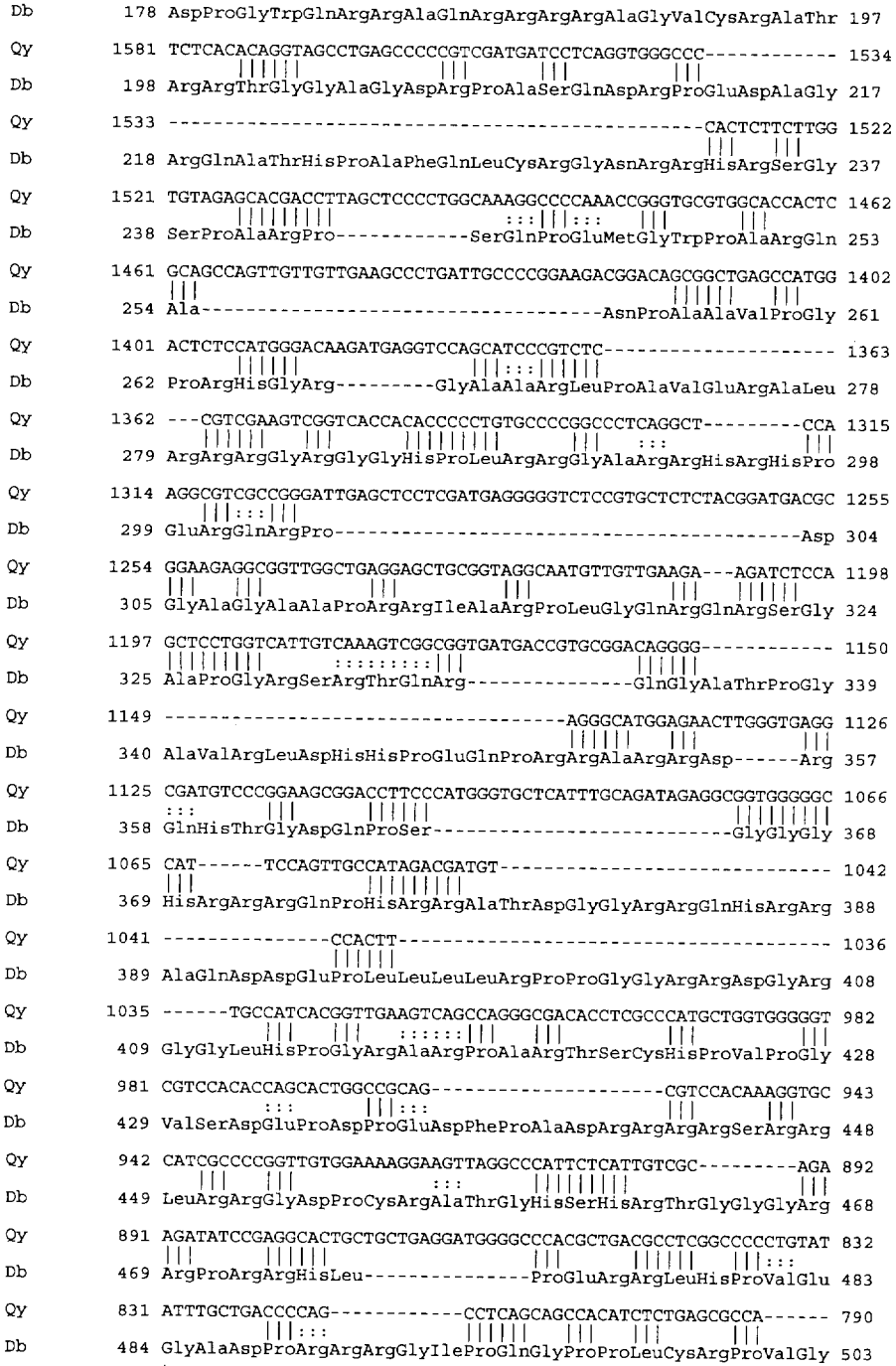
Alignment Scores:

Pred. No.:	6.97e-08	Length:	798
Score:	210.50	Matches:	177
Percent Similarity:	30.42%	Conservative:	49
Best Local Similarity:	23.82%	Mismatches:	239
Query Match:	5.19%	Indels:	278
DB:	4	Gaps:	40

US-09-914-958B-35 (1-2177) x US-09-252-991A-23774 (1-798)

```
Qy         52 GCGCCGGCGGCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCG 111
              |||   |||       |||   ||||||||||||                       |||
Db          3 AlaAlaAlaMetSerAlaGlyAlaSerAlaArg---------------ArgCysTrpPro 17

Qy        112 AGAGCGGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGC 171
                ||||||||||   :::       :::           |||   |||   |||
Db         18 MetAlaGlyGlySerMetAlaCysValAlaArgTrpThrPro---CysArgArgGlySer 36

Qy        172 TGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTG 231
                |||
Db         37 Cys------------------------------------------------------- 37

Qy        232 CAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATG 291

Db         37 ------------------------------------------------------------ 37

Qy        292 GTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACA 351
                           |||      :::      |||||||||
Db         38 ---------------ArgTrpLeuAlaIleCysArgLeuArgSerVal------------ 48

Qy        352 ATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGG 411
                                             ||||||||||||
Db         49 ---------------------------------ProArgSerGly--------------- 52

Qy        412 TCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGC---------AGGGGAACG 462
                     |||:::   ||||||   |||                          |||
Db         53 ------ArgAlaSerArgSerThrThrTrpCysThrAlaProProProValSerThrThr 70

Qy        463 CCATTGGGGTCACAGCCTGCGACATCGACGGGGACG-----------GCCGGGAGGAGA 510
                 |||   :::   |||   |||||||||                  ||||||   |||
Db         71 ArgLeuAlaThrValProProThrSerThrAlaCysAlaMetCysTrpAlaGlyCysArg 90

Qy        511 TCTACTTCCTCAACACCA------------ATAATGCCTTCTCGGGGGTGGCCACGTACA 558
                 :::|||   |||                    |||:::   |||      |||
Db         91 ProAlaAlaSerGlyProGlyAlaCysCysSerSerProAlaProGlySerThrArgArg 110

Qy        559 CCGACAAGTTGTTCAAGTTCC------------------------GCAATAACCGGTGGG 594
                 |||   :::       |||:::                          :::|||   |||
Db        111 ProMetAlaAlaGlySerThrArgSerArgArgArgPheProArgSerIleProAspGly 130

Qy        595 AAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC 654
                    |||         ||||||   |||         :::         ||||||
Db        131 SerCysSerThrProSerArgSerArgTrpThrAlaGlyTyr------ArgLeuProAla 148

Qy        655 GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATT 714
                    ||||||      |||   |||
Db        149 CysAlaTrpProAlaSerThrValArgGly----------------------------- 158

Qy        715 ACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACC 774

Db        158 ------------------------------------------------------------ 158

Qy        775 TCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCA---GCAAAT 831
                                                                |||||      :::
Db        159 ---------------------------------------------GlySerGlyCysSer 163
```

```
Qy       832 ATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCT 891
Db       164 ThrArgCysValArgGlyThrGlyTrpSerAlaSerArgProCysMetProThrGlySer 183

Qy       892 TCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTG 951
Db       184 ThrProThrMetArgProGly-----------CysTrpProSerCysCysAlaProMet 199

Qy       952 TGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCC 1011
Db       200 ProAlaAlaArgArg---------TrpArgThrAlaThrSerValSerThrThrArgPro 216

Qy      1012 TGG-----------CTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACT 1059
Db       217 TrpArgCysThrArgTrpSerThrThrCysAlaSerAlaTrpAlaSer------AlaAsn 234

Qy      1060 GGA-----------------ATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATG 1101
Db       235 GlyProThrAsnIleArgTyrAlaAlaProAlaAlaSerValAlaAlaThrAlaAlaPro 254

Qy      1102 GGAAGGTCCGCT-----------------------------------------------TCC 1116
Db       255 GlyArgSerAlaGlyCysArgAlaIleProAlaThrAlaArgAlaThrArgArgCysSer 274

Qy      1117 GGGACATCGCCT-------------------------------------CACCCAAGT 1137
Db       275 GlyThrProAspGlyCysGlnProValValGlyAlaSerLeuAlaAlaTrpArgGlyAla 294

Qy      1138 TCTCCATGC----------------------CCTCCC-----------CTGTCCGCA 1161
Db       295 ValProCysGlyLeuLeuArgProGlyValArgProProGlnProPheGlyLeuArgHis 314

Qy      1162 CGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTG 1221
Db       315 ArgLeuAspProPro----------------------ArgArgArgArgLeuProVal 326

Qy      1222 CCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACC 1281
Db       327 ProArgAlaAla------ProCysAlaAlaGly-ArgHisPro------ValAlaAspGl 342

Qy      1282 CCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGG 1341
Db       342 uProGlyArgAlaThrTyrArgAlaArgArgIleArgAla---------------AlaGl 357

Qy      1342 GTGTGGTGACCGACTTCGACGG-----------AGACGGGATGCTGGACCTCATCTTGT 1389
Db       357 yLeuProAspAlaLeuCysArgGlyAsnAlaProAlaArgProAlaArgProGlnThrVa 377

Qy      1390 CCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGT-----------------CTTCCGGG 1431
Db       377 lAlaAlaArgLeuArgArgThrGluProGlyArgArgArgGlnArgGlyGlyTrpProGl 397

Qy      1432 GCAATCAGGG-----------CTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCC 1479
Db       397 yValAlaGlyArgGlyPheProAlaGlnArgArgIle---------GlyValGlyLysPr 414

Qy      1480 GGTTTGGGGCCTTTGCCAGGGGAGC------TAAGGTCGTGCTCTACACCAAGAAGAGTG 1533
Db       414 oProAlaGlyLeuAlaArgThrGlyLeuArgAlaSerArgArgLeuAlaProGlyGlyAr 434

Qy      1534 GGGCCCACCT------GAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587
Db       434 gGlyAlaProArgArgArgAsnHisArgThrProAlaArgLeuProGlySerAlaCysTy 454

Qy      1588 CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGA------GGTGACGT 1641
Db       454 rArgGlyAla---------------------------ArgSerGlyProThrGlyAlaVa 465

Qy      1642 GGCCAGATGGCAAGATGGTGAGCCGGAACGTGGC-----------------CAGCGGGG 1683
Db       465 lAlaAlaAlaProProAlaProProAspArgGlyPheProAlaCysLeuArgArgAlaAl 485

Qy      1684 AGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAG 1743

                                                                  ::: ::
Db       485 aAlaTyrLeuProThrAlaMetProGlyAlaProGlyGlnAlaProValAlaGlyArgTh 505

Qy      1744 CCCCACTGGA---------GTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCA 1794
Db       505 rAlaAlaGlyGlyThrArgProAlaProArgLeuLeuArgProGlyAlaLeu-------- 522

Qy      1795 TGGACACCAATGAATGCATCCAGTTCCCATT--------------------CGTGTGCC 1833
Db       523 ------------LeuHisAspLeuProAlaPheHisArgGlyHisSerGlyAlaValPr 538

Qy      1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 1893
Db       538 oAlaGlnArg------------GlnAspLeuSerPhePheGlnTyrGlyAlaArgArgAl 554

Qy      1894 AGTGCAGTCG---GGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGA 1950
Db       554 aAlaGlnThrGlyGlyLeuArgValSerArgGlyLeuHisValArgArg------ValAr 572

Qy      1951 GCCCTGT 1957
Db       572 gProCys 574

RESULT 13
US-09-252-991A-16655
; Sequence 16655, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16655
;   LENGTH: 977
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16655

Alignment Scores:
Pred. No.:                8.31e-08      Length:          977
Score:                    210.00        Matches:         226
Percent Similarity:       28.05%        Conservative:    36
Best Local Similarity:    24.20%        Mismatches:      253
Query Match:              5.08%         Indels:          420
DB:                       4             Gaps:            51

US-09-914-958B-35 (1-2177) x US-09-252-991A-16655 (1-977)

Qy      1845 GCTTGTCTCGAGGGCACACGAATGGGAACTGGATGCATTCATTGGTGTCCATGCAATGGC 1786
Db       121 AlaArgLeuAlaGlyGlnArgGlyAlaThrAlaAlaAla---------------------- 133

Qy      1785 CATTTTCCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTG------GGTCCTGAA 1732
Db       134 ------ProAlaArgArg-----------ThrProGlyGlyLeuProArgGlyAlaPro 147

Qy      1731 GTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGGCCA 1672
Db       148 AlaArgProArgProGlyHisArg------ValAlaValArgAlaGlyGlnArgSerArg 165

Qy      1671 CGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCT 1612
Db       166 ArgThrArgProPro-----------------ArgGlyProArgCysLeu------Pro 177

Qy      1611 TCCCCAGGC----------------------------CAAAGTGTGCCACGGGCTCCA 1582
```


Db 178 AspProGlyTrpGlnArgArgAlaGlnArgArgArgArgAlaGlyValCysArgAlaThr 197
Qy 1581 TCTCACACAGGTAGCCTGAGCCCCCGTCGATGATCCTCAGGTGGGCCC---------- 1534
Db 198 ArgArgThrGlyGlyAlaGlyAspArgProAlaSerGlnAspArgProGluAspAlaGly 217
Qy 1533 ----------CACTCTTCTTGG 1522
Db 218 ArgGlnAlaThrHisProAlaPheGlnLeuCysArgGlyAsnArgArgHisArgSerGly 237
Qy 1521 TGTAGAGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTC 1462
Db 238 SerProAlaArgPro----------SerGlnProGluMetGlyTrpProAlaArgGln 253
Qy 1461 GCAGCCAGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGG 1402
Db 254 Ala----------AsnProAlaAlaValProGly 261
Qy 1401 ACTCTCCATGGGACAAGATGAGGTCCAGCATCCCGTCTC---------- 1363
Db 262 ProArgHisGlyArg---------GlyAlaAlaArgLeuProAlaValGluArgAlaLeu 278
Qy 1362 ---CGTCGAAGTCGGTCACCACACCCCCTGTGCCCCGGCCCTCAGGCT---------CCA 1315
Db 279 ArgArgArgGlyArgGlyGlyHisProLeuArgArgGlyAlaArgArgHisArgHisPro 298
Qy 1314 AGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGATGACGC 1255
Db 299 GluArgGlnArgPro----------Asp 304
Qy 1254 GGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGA---AGATCTCCA 1198
Db 305 GlyAlaGlyAlaAlaProArgArgIleAlaArgProLeuGlyGlnArgGlnArgSerGly 324
Qy 1197 GCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAGGGG---------- 1150
Db 325 AlaProGlyArgSerArgThrGlnArg----------GlnGlyAlaThrProGly 339
Qy 1149 ----------AGGGCATGGAGAACTTGGGTGAGG 1126
Db 340 AlaValArgLeuAspHisHisProGluGlnProArgArgAlaArgArgAsp------Arg 357
Qy 1125 CGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGCAGATAGAGGCGGTGGGGGC 1066
Db 358 GlnHisThrGlyAspGlnProSer----------GlyGlyGly 368
Qy 1065 CAT------TCCAGTTGCCATAGACGATGT---------- 1042
Db 369 HisArgArgArgGlnProHisArgArgAlaThrAspGlyGlyArgArgGlnHisArgArg 388
Qy 1041 ----------CCACTT---------- 1036
Db 389 AlaGlnAspAspGluProLeuLeuLeuLeuArgProProGlyGlyArgArgAspGlyArg 408
Qy 1035 ------TGCCATCACGGTTGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGT 982
Db 409 GlyGlyLeuHisProGlyArgAlaArgProAlaArgThrSerCysHisProValProGly 428
Qy 981 CGTCCACACCAGCACTGGCCGCAG----------CGTCCACAAAGGTGC 943
Db 429 ValSerAspGluProAspProGluAspPheProAlaAspArgArgArgArgSerArgArg 448
Qy 942 CATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGC---------AGA 892
Db 449 LeuArgArgGlyAspProCysArgAlaThrGlyHisSerHisArgThrGlyGlyGlyArg 468
Qy 891 AGATATCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTAT 832
Db 469 ArgProArgArgHisLeu----------ProGluArgArgLeuHisProValGlu 483
Qy 831 ATTTGCTGACCCCAG----------CCTCAGCAGCCACATCTCTGAGCGCCA------ 790
Db 484 GlyAlaAspProArgArgArgGlyIleProGlnGlyProProLeuCysArgProValGly 503
Qy 789 ---GAATGCCCCGGGAGAGGT----------CACTGGCCTCAGGGTCCATTTCAATGA 745
Db 504 AlaGlyHisProGlyThrGlyAlaGluHisArgHisArgProHisArgGlyMetGluGly 523
Qy 744 GGGCAT---CAGGGCCCACATTACCGTAGGCGT----------AATTGGCAA 706
Db 524 ArgHisArgArgProProHisGlnArgArgArgArgAlaAlaGluGluThrArgGlyArg 543
Qy 705 TGTAGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAA 646
Db 544 CysArgProGlyLeuGlyGluAspProArgArgGlnLysArgGlyGlyArgThrArgArg 563
Qy 645 AGAGGCTGGCCACACCACGGGCCACGTTGACCT---------- 613
Db 564 ArgArgGlnProAlaHisArgValArgAlaProAlaAlaGlyArgArgLeuAlaGluArg 583
Qy 613 ---------- 613
Db 584 ArgAlaThrAspProAlaAlaGlyArgGlnGlyAspLeuLeuHisArgGlyAlaGlyAla 603
Qy 612 ----------CATCGCTCAGGATGTCTT 595
Db 604 ArgGlnProAlaGlnAlaProGlyGlyGlyArgArgArgLeuHisArgProGly------ 621
Qy 594 CCCACCGGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGG 535
Db 622 ---AlaGlyTyr----------ArgLeuProGlnAlaArgArg 631
Qy 534 CATTATTGGTGTTGAGGAAGTAGATCT------CCTCCCGGCCGTCCCCGTCGATGT--- 484
Db 632 ArgGlyGlyGlyGlySerAlaThrAlaHisProAlaGlyLeuArgArgArgThrAsp 651
Qy 483 ----------CGCAGGCTGTGACCCCAATGGCGTTCC 457
Db 652 GlnAlaGlyGlyProGlyLeuAlaGlnAlaGlyArgArgAlaValProArg---AlaGln 670
Qy 456 CCTGCCGGTCCC----------GCAGCGCGT 436
Db 671 ProAlaGlyProGluArgGluArgArgAlaGlyProArgArgArgArgArgGlyAlaArg 690
Qy 435 AGTAGGGTGAGCTGCGCTCATCGACCG---------- 409
Db 691 AspArg----------ArgArgProGlyThrGlyGlyGlyArgProGlnAlaAlaGln 706
Qy 408 ---CGATGTTCACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGG------ 358
Db 707 ArgArgLeuGluProGlyLysProGlyProGlyHis---GluArgProGlyGlyGluGly 725
Qy 357 ----------GTCCATTGTACCCCGCCACGACGATCT--------- 331
Db 726 ArgArgProValProHisLeuAspAlaGlnCysLeuGlyHisArgArgSerArgArgArg 745
Qy 330 ---------CAAAGTCCCCATCATGGTCCA----------CAT 307
Db 746 AlaAspAlaArgAlaProGlyHisGlyProGlyArgAspGlyArgArgAlaAspArgArg 765
Qy 306 CAGTAACTGCCA----------CACCATAGTTGAGCT 280
Db 766 GlnAlaSerProValArgProGlyGlyAspProArgGlyValLeuHisArgSerGlySer 785
Qy 279 GGGTGGGATTACTGTCATA---------- 261
Db 786 -GlyGlyArgArgValValProGlyAlaGlyGluGlyCysArgProGlyLeuProGlyGl 805
Qy 260 ----------GTCAGGAGGCAGAACTGAGTTGGTGAC---------- 234
Db 805 yGluLeuProValArgArgGlnArgProArgHisAspProGlyGlyGlnArgArgLeuAr 825
Qy 234 ---------- 234
Db 825 gProArgGlyGlyAlaSerArgGlnProProGlyArgArgLeuAlaGlyGlyGlyGlnGl 845

```
Qy       233 -------TGCAGTGAACATGGGTTCAGCCCGCTGGGACCCCTCAGTGATGGGCAGAAACC 181
Db       845 yGlyLeuGlyThrValHisGlyLeuArgProValAla------GlyAspGly---ArgPr 862

Qy       180 AGAGCAGCAGCAGGAACGGTAACATCCTGGACATGCCGGGGT-CAGCGCTCGGAGCCATC 122
Db       862 oProGlyArgHisArgArgHisHisProArgProSerAspProArgArgSerGlyProGl 882

Qy       121 CTCCCGCTCTCGGCCCCGCCGCCTAGGGGCGTGGGAAGCG-------------------- 82
Db       882 y------SerArgProAlaArgAlaGlyThrArgProAlaHisLeuIleProAlaAsnLe 900

Qy        81 -------------------GGCGCTCGCTGCCGCCTCTGCCGCCGGCGCCGCCGCCTG-- 43
Db       900 uGlySerArgProThrThrGlyAlaHisTyrProLeuAlaArgArgAlaArgGlyLeuPh 920

Qy        42 ------------------------------------------------------------CT 41
Db       920 eSerTrpProArgThrGlySerSerAlaGlyLeuThrAlaLeuArgAspProProProLe 940

Qy        40 TGCTCCCAGCCCCGGTCCCGGGCTGGCCTCGAGCCTCCCG 1
Db       940 uThrProSerProAlaThrHisAlaAlaSerSerProPro 953
```

RESULT 14
US-09-252-991A-30843
; Sequence 30843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Alignment Scores:

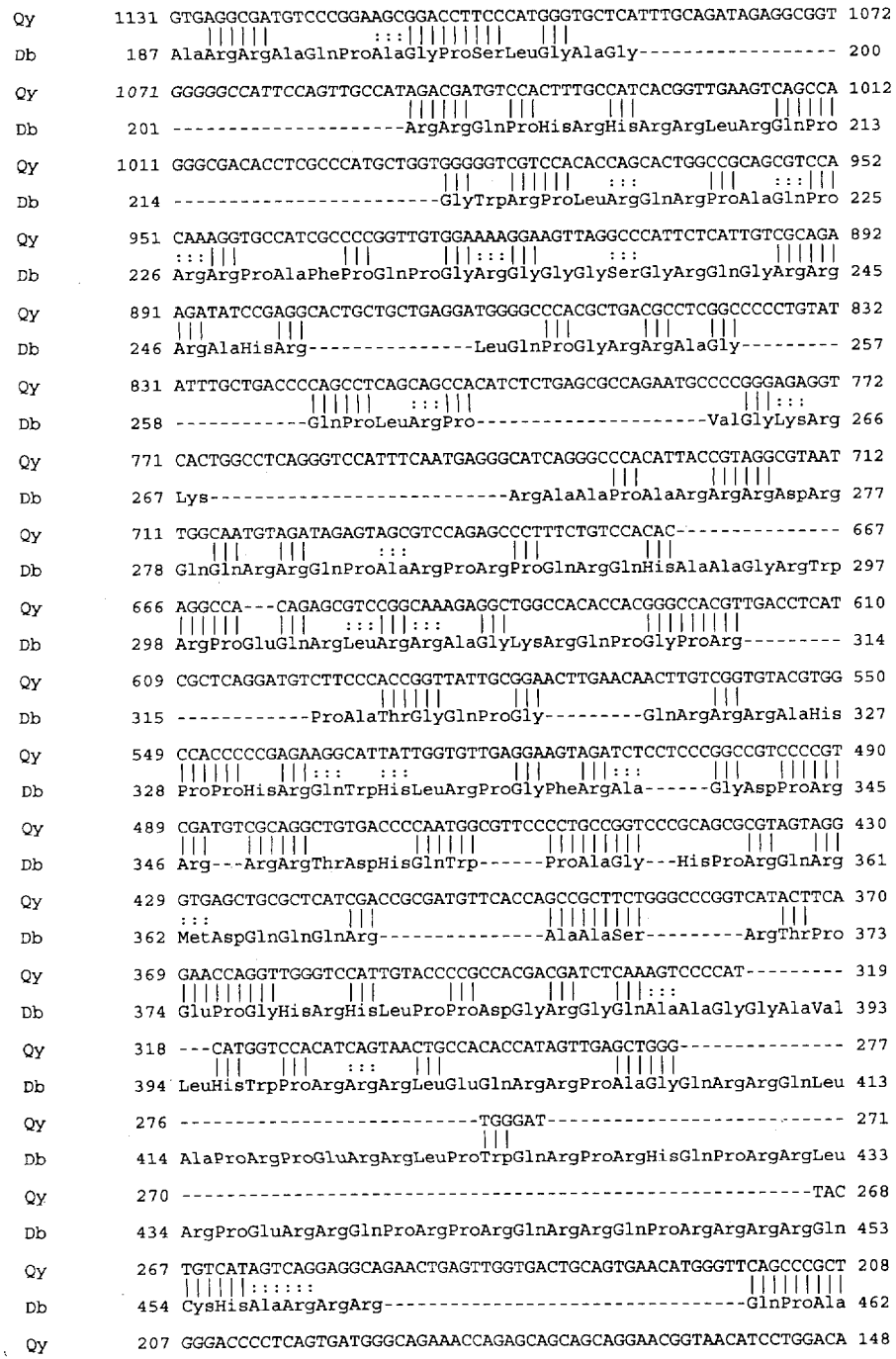
Pred. No.:	1.2e-07	Length:	663
Score:	207.00	Matches:	183
Percent Similarity:	30.25%	Conservative:	56
Best Local Similarity:	23.16%	Mismatches:	203
Query Match:	5.01%	Indels:	348
DB:	4	Gaps:	44

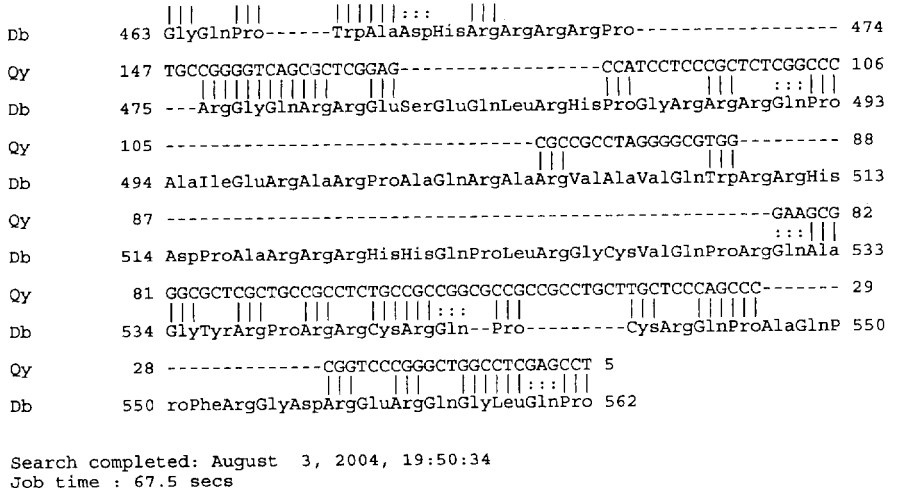
US-09-914-958B-35 (1-2177) x US-09-252-991A-30843 (1-663)

```
Qy      1950 TCCACCAGCCCACGCAGGCTGTGCCATCCTCGTTGGGCTCGTAGCCCCGACTGCACTTCT 1891
Db        62 SerThrSerArgArg-----------ProGlyTrpSerProAlaValProAlaGlyAla 77

Qy      1890 TGT-----------------------------------------------TGGTCCGGCACC 1876
Db        78 CysCysTrpLysThrArgAsnValAlaGlyArgSerAlaThrAspAlaTrpSerSerAla 97

Qy      1875 TGTAGCTTCCATAGGTGTTGACACATACGGGCTTGTCTCGAGGGCACACGAATGGGAACT 1816
Db        98 ProAlaProAlaAsnCys-----------CysCysLeu---------------------- 106

Qy      1815 GGATGCATTCATTGGTGTCCATGCAATGGCCATTTTCCTGCTGGGAGAATCCTTGGCCAC 1756
Db       107 --------------------------------PheProAlaGly--------------- 110

Qy      1755 ACTCCAGTGGGGCTGGGTCCTGAAGTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCA 1696
```

```
Db       111 ----------------------------ProSerPro---------AlaSerProAla 117

Qy      1695 CTGAGTTCATCTCCCCGCTGGCCACGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCA 1636
Db       118 ProAlaAlaCysArgArgTrpProArgAla---------AlaCysHis------------ 130

Qy      1635 CCTCCACACTGCTGGCTTCATCCTTCCCCAGGCCAAAGTGTGCCACGGGCTCCATCTCAC 1576
Db       131 ------------Trp---------------------------------------------- 131

Qy      1575 ACAGGTAGCCTGAGCCCCCGTCGATGATCCTCAGGTGGGCCCCACTCTTCTTGGTGTAGA 1516
Db       132 ------------------------------------ProAlaSerAlaTrpTrp--- 137

Qy      1515 GCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGC------------GTGGCA 1468
Db       138 -----------------TrpLeuAlaProAlaArgCysCysTrpProAlaProProAla 151

Qy      1467 CCACTCGCAGCC------AGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCG 1414
Db       152 ProAlaSerAlaGlyArgAlaCysCysAlaSerProAsn------------ArgArgArg 167

Qy      1413 GCTGAGCCATGG------ACTCTCCATGGGACAAGATGAGGTCCAGCATCCCGTCTCCGT 1360
Db       168 ArgGluProTrpProProSerProTyrAlaSerArgAlaGlyProAlaSerCysGlyArg 187

Qy      1359 CGA-----------------AGTCGGTCACCACACCCCCTGTGCCCCGGCCCCTCAGGCT 1318
Db       188 ProProAlaCysSerProValAlaThrAlaProThrAlaThrCysSerProProSerAla 207

Qy      1317 CCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGATGA 1258
Db       207 ------------------------------------------------------------ 207

Qy      1257 CGCGGAAGAGGCGGTTGGCTGAGG---AGCTGCGGTAGGCAATGTTGTTGAAGAAGATCT 1201
Db       208 ------ArgSerAlaTrpLysProCysAlaCysAlaLysAlaValGlySerAlaArgSer 225

Qy      1200 CCAGCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAGGGGAGGGCATGG 1141
Db       226 ProAla----------------------------------------------AsnAlaTrp 230

Qy      1140 ---------AGAACTTGGGTGAGGCGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCA 1090
Db       231 ProAlaAlaSerAlaTrpCysProThrCysAsnSerAlaArgProSerAlaIleAlaSer 250

Qy      1089 TTTGCAGATAGA------GGCGGTGGGGGCCATTCCAGTTGCCATAGACGATGTCCACTT 1036
Db       251 ThrAlaArgArgSerProSerThrAsnGlyArg-ProAlaAlaCys-ProThrThrThrP 270

Qy      1035 TGCCATCACGGTTGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGTCGTCCA 976
Db       270 roProAlaAsnAlaProAlaSerAlaAlaAlaAsnTrpPro---TrpTrpLysAlaProS 289

Qy       975 CACCAGCA----------------------------CTGGCCGCA---------------- 959
Db       289 erProAlaMetProProSerThrSerAlaThrArgLeuAlaAlaCysGlyHisAspGlyA 309

Qy       958 ----------------------------------GCGTCCACAAAGGTGCCATCGCCCC 934
Db       309 spAlaGlyArgAlaSerGlnThrProTrpArgAspThrSerProCysValProAsnCysV 329

Qy       933 GGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATATCCGAGGCACTGC 874
Db       329 alAsnTrpArgArgArgIle---ProTrpSerAlaValAlaLysThrSerPro------- 345

Qy       873 TGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTATATTTGCTGACCCCAGCCT 814
Db       346 ----------------------TrpProArgTrpProGlyMetProAlaGlyProArgP 358

Qy       813 CAGCAGCCACATCTCTGAGCGCCAGAATGCCCCGGGAGAGGTCACTGGCCTCAGGGTCCA 754
```

```
Db        358 roSerCysThrAlaValAlaAla-TrpAlaProAlaArgGlyGlyTyrAlaAlaAlaPro 377

Qy        753 TTTCAATGAGGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAATGTAGATAGAGT 694

Db        378 ArgAsnSerCysSerAlaGly---ArgLeuArgArgArg-ValArgArgSerAlaArgAr 396

Qy        693 AGCGTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAG---------------------- 656

Db        396 gAlaTrpArg-ProTrpProAlaGlyArgAlaThrProAlaSerAspAlaArgArgAspS 416

Qy        655 --------------CGTCCGGCAAAGAGGCTGGCCACACCACGGGCCACGTTGACCTCAT 610

Db        416 erAlaSerArgGlyArgProAlaThrAlaAlaAspHisProAlaAlaTrpValSerAlaA 436

Qy        609 CGCTCAGGATGTCTTCCCACCGGTTATTGC------------GGAACTTGAACAACTTGT 562

Db        436 laArgArgThrSerSerAlaProIle-AlaGlySerAlaProGlySerGlyThrAlaPro 455

Qy        561 CGGTGT---------------------------------------------------- 556

Db        456 ArgCysHisProValArgLysAspGlyAlaGlyProAlaSerThrGlyArgSerArgArg 475

Qy        556 ---------------------------------------------------------- 556

Db        476 ArgTrpAlaAlaAlaProAlaArgArgAlaArgThrGlyIleArgArgThrAlaAlaAla 495

Qy        555 ---ACGTGGCCACCCCCGAGAAGGCATTATTGGTGTTGAGGAAGTAGATCTCCTCCCGGC 499

Db        496 AlaCysTyrProProProAlaPro--------------GlyArgArgThrProAlaAla 510

Qy        498 CGTCCCCGTCGATGTCGCAGGCTGTGACCCCAATGGCGTTCCCCTGCCGGTCCCGCAGCG 439

Db        511 Arg---ArgArgSerAlaArgThrAlaProArg---------ProAlaIlePro------ 524

Qy        438 CGTAGTAGGGTGAGCTGCGCTCATCGACCGCGATGTTCA---------------------- 400

Db        525 -----------------AlaHisArgProSerAlaSerAlaThrGlyTrpProGlyPro 538

Qy        399 ------------CCAGCCGCTTCTGGG------------CCCGGTCATACTTCAGAACCA 364

Db        539 AlaLeuAlaCysProAlaAlaAlaGlyArgArgProSerProAlaProAlaProAlaPro 558

Qy        363 GGTTGGGTCCATTGTACCCCG------CCACGACGA------TCTCAAAGTCCCCATCAT 316

Db        559 Ala---ProHisAlaArgProArgAlaProArgArgArgProGlyArgSerProAlaThr 577

Qy        315 GGTCCACATCAGTAACTGCCACACCATAGTTGAGCTGGGTGGGATTACTGTCATAGTCAG 256

Db        578 AlaProAla------------------------------------------------- 580

Qy        255 GAGGCAGAACTGAGTTGGTGACTGCAGTGAACATGGGTTCAGCCCGCTGGG---ACCCCT 199

Db        581 ------------------------------AlaPheAlaAlaProAlaGlyProThrPro 590

Qy        198 CAGTGATGGGCAGAAACCAGAGCAGCAGCAGGAACGGTAACATCCTGGACATGCCGGGGT 139

Db        591 ---------AlaArgProProAlaAlaAla-----------------------ThrGly 599

Qy        138 CAGCGCTCGGAGCCATCCTCCCGCTCTCGGCCCCGCCGCCTAGGGGCGTGGGAAGCGGGC 79

Db        600 ArgProSerAlaProProArgProAlaArgArgArgValArgTrpProTrpArgSer--- 618

Qy         78 GCTCGCTGCCGCCTCTGCCGCCGGCGCCGCCGCCTGCTTGCTCCCAGCCCCGGTCCCGGG 19

Db        619 AlaArgCys--------------SerProArgArgProThrProArgSerAlaProGly 633

Qy         18 CTGGCCTCGAGCCTCCCG 1

Db        634 AlaAlaThrGlyAlaPro 639

RESULT 15
US-09-252-991A-21920
; Sequence 21920, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 21920
;   LENGTH: 822
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21920

Alignment Scores:
Pred. No.:                2.43e-07        Length:          822
Score:                    203.50          Matches:         177
Percent Similarity:       30.40%          Conservative:    34
Best Local Similarity:    25.50%          Mismatches:      204
Query Match:              4.92%           Indels:          279
DB:                       4               Gaps:            40

US-09-914-958B-35 (1-2177) x US-09-252-991A-21920 (1-822)

Qy       1719 CCCGGGGGTAGAGGATCTCCAGCACTGAGTTCA---TCTCCCCGCTGGCCACGT------ 1669

Db         27 ProGlyGlnArgArgArgProAlaAlaAlaGlnArgGlnArgArgGlnProArgArgGln 46

Qy       1668 TCCGGCTCACCATCTTGC---------------------------------------- 1651

Db         47 AlaGlyGlnProGlyAlaAlaGlyAspHisHisArgGlnProArgGlnGlnArgLeuArg 66

Qy       1650 CATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCTTCCCCAGGCCAAAGTGTGCCA 1591

Db         67 HisProArgGlnProGlyArgHis---------GluProAlaProGlyArgArgArgPro 83

Qy       1590 CGGGCTCCATCTCACACAGGTAGCCTGAGCCCCCGTCGATGATCCTCAGGTGGGCCCCAC 1531

Db         84 AlaGlnProAlaGlyArgProAspLeuGlnProGlyArgArgProArgGlyAlaGlyArg 103

Qy       1530 TCTTCTTGGTGTAGAGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGCGTG 1471

Db        104 Gln---------------------------ProGlyGlnProProGlyHis--------- 111

Qy       1470 GCACCACTCGCAGCCAGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCGGCT 1411

Db        112 AlaPro-----------------------------GlyProGlyArgGlnProAlaAla 121

Qy       1410 GAGCCATGGACTCTCCATGGGACAAGATGAGGTCCAGCATCCCGTCTC------CGTCGA 1357

Db        122 TyrArgTrp---------------ArgAlaGlyGlnProGlyArgProProGlyGlnPro 136

Qy       1356 AGTCGGTCACCACACCCCCTGTGCCCCGGCCCTCAGGCTCCAAGGCGTCGCCGGGATTGA 1297

Db        137 GlyArgGlnProArgProAlaGluArgGlnProArgGlnArgArgArgArgArg------ 154

Qy       1296 GCTCCTCGATGAGGGGGTCTC---------------CGTGCTCTCTACGGATGACGCGGA 1252

Db        155 AlaGlnGlnArgGlnGlyLeuAlaAspAlaGlyHisArgAlaValArgGlnGlnArg--- 173

Qy       1251 AGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGAAGATCTCCAGCTCCT 1192

Db        174 -----------------------------------------ArgArgHisProGlyAla 179

Qy       1191 GGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAGGGGAGGGCATGGAGAACTTGG 1132

Db        180 ValAlaGlyAspSerArgArg---------------------------------------- 186
```





Search completed: August 3, 2004, 19:50:34
Job time : 67.5 secs

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:48:01 ; Search time 108 Seconds
(without alignments)
12646.069 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 2582470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161654_29816/app_query.fasta_1.2375
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09914958_@CGN_1_1_15_@runat_03082004_161654_29816
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3292	81.2	618	12	US-10-332-929A-20	Sequence 20, Appl
2	3190.5	78.7	653	12	US-10-332-929A-2	Sequence 2, Appli
3	3190.5	78.7	653	15	US-10-295-027-183	Sequence 183, App
4	3190.5	78.7	653	16	US-10-188-832-58	Sequence 58, Appl
5	3181.5	78.4	637	12	US-10-332-929A-19	Sequence 19, Appl
6	3179	78.4	612	12	US-10-332-929A-21	Sequence 21, Appl
7	2337	57.6	1597	15	US-10-295-027-185	Sequence 185, App
8	2337	57.6	1597	16	US-10-188-832-60	Sequence 60, Appl
9	1967.5	48.5	418	15	US-10-295-027-181	Sequence 181, App
10	1967.5	48.5	418	16	US-10-188-832-56	Sequence 56, Appl
11	1492	36.8	306	10	US-09-892-877-247	Sequence 247, App
12	1492	36.8	306	10	US-09-948-783-249	Sequence 249, App
c 13	318	7.7	19662	15	US-10-084-846A-6	Sequence 6, Appli
14	310	7.6	19723	15	US-10-084-846A-5	Sequence 5, Appli
c 15	291	7.0	19723	15	US-10-084-846A-5	Sequence 5, Appli
c 16	286.5	6.9	19725	15	US-10-084-846A-4	Sequence 4, Appli
c 17	284	6.9	19608	15	US-10-084-846A-8	Sequence 8, Appli
18	283	7.0	19608	15	US-10-084-846A-8	Sequence 8, Appli
19	276.5	6.8	437	16	US-10-437-963-132079	Sequence 132079,
c 20	273	6.6	19695	15	US-10-084-846A-3	Sequence 3, Appli
21	272.5	6.7	19695	15	US-10-084-846A-3	Sequence 3, Appli
22	267	6.6	50	10	US-09-948-783-125	Sequence 125, App
23	267	6.6	51	10	US-09-892-877-124	Sequence 124, App
c 24	259	6.3	19652	15	US-10-084-846A-7	Sequence 7, Appli
25	257	6.3	1367	9	US-09-801-368-108	Sequence 108, App
26	255	6.3	19725	15	US-10-084-846A-4	Sequence 4, Appli
27	254	6.3	427	16	US-10-437-963-199493	Sequence 199493,
28	253.5	6.2	5179	9	US-09-922-217-1068	Sequence 1068, Ap
29	253.5	6.2	5179	9	US-09-833-263-1068	Sequence 1068, Ap
30	253.5	6.2	5179	13	US-10-025-380-1068	Sequence 1068, Ap
31	250.5	6.2	19662	15	US-10-084-846A-6	Sequence 6, Appli
32	249	6.1	528	10	US-09-840-746-20	Sequence 20, Appl
33	244	6.0	19652	15	US-10-084-846A-7	Sequence 7, Appli
34	243.5	6.0	497	16	US-10-437-963-125004	Sequence 125004,
35	238.5	5.9	487	16	US-10-437-963-137896	Sequence 137896,
36	237.5	5.9	377	16	US-10-437-963-133141	Sequence 133141,
37	229.5	5.7	738	9	US-09-978-979-6	Sequence 6, Appli
38	229.5	5.7	738	14	US-10-057-487-6	Sequence 6, Appli
39	224	5.5	1259	14	US-10-260-715-8	Sequence 8, Appli
40	223	5.5	417	16	US-10-437-963-143835	Sequence 143835,
41	215.5	5.3	613	15	US-10-260-937-16	Sequence 16, Appl
42	215.5	5.3	2080	15	US-10-353-690-36	Sequence 36, Appl
43	213	5.3	465	16	US-10-437-963-174113	Sequence 174113,
44	212	5.2	36	10	US-09-892-877-255	Sequence 255, App
45	212	5.2	36	10	US-09-948-783-257	Sequence 257, App

ALIGNMENTS

RESULT 1
US-10-332-929A-20
; Sequence 20, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-332-929A-20

Alignment Scores:

Pred. No.:	8.57e-209	Length:	618
Score:	3292.00	Matches:	618
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	81.16%	Indels:	0

```
DB:                 12          Gaps:          0

US-09-914-958B-35 (1-2177) x US-10-332-929A-20 (1-618)

Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer  20

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp  40

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp  60

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla  80

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy      1587 CCCGTGGCACACCTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyTrp 600

Qy      1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 TrpSerProValLeuLysIleValThrProGlnValGlyLysSerLeuGlyPro 618

RESULT 2
US-10-332-929A-2
; Sequence 2, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 653
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-332-929A-2
```

```
Alignment Scores:
Pred. No.:                4.35e-202      Length:          653
Score:                    3190.50        Matches:         603
Percent Similarity:       97.90%         Conservative:    2
Best Local Similarity:    97.57%         Mismatches:      2
Query Match:              78.66%         Indels:          11
DB:                       12             Gaps:            1

US-09-914-958B-35 (1-2177) x US-10-332-929A-2 (1-653)

Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy       327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy      1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                                 :::|||:::|||   ||||||
Db       600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607

RESULT 3
US-10-295-027-183
; Sequence 183, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
```

```
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 183
;   LENGTH: 653
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-183

Alignment Scores:
Pred. No.:                4.35e-202       Length:          653
Score:                    3190.50         Matches:         603
Percent Similarity:       97.90%          Conservative:    2
Best Local Similarity:    97.57%          Mismatches:      2
Query Match:              78.66%          Indels:          11
DB:                       15              Gaps:            1

US-09-914-958B-35 (1-2177) x US-10-295-027-183 (1-653)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500
```

```
Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
Db        600 ----------------------------ThrLeuGlyGlnSerProGlyPro 607

RESULT 4
US-10-188-832-58
; Sequence 58, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
;  APPLICANT: Mack, David H.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
;  TITLE OF INVENTION:  and Methods of Screening for Modulators of Bladder
;  TITLE OF INVENTION:  Cancer
;  FILE REFERENCE: 018501-002330US
;  CURRENT APPLICATION NUMBER: US/10/188,832
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US 60/302,814
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: US 60/310,099
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/343,705
;  PRIOR FILING DATE: 2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/372,246
;  PRIOR FILING DATE: 2002-04-12
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 653
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-188-832-58

Alignment Scores:
Pred. No.:              4.35e-202       Length:         653
Score:                  3190.50         Matches:        603
Percent Similarity:     97.90%          Conservative:   2
Best Local Similarity:  97.57%          Mismatches:     2
Query Match:            78.66%          Indels:         11
DB:                     16              Gaps:           1

US-09-914-958B-35 (1-2177) x US-10-188-832-58 (1-653)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400
```

```
Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                                :::|||:::|||   ||||||
Db        600 ------------------------------ThrLeuGlyGlnSerProGlyPro 607
```

RESULT 5
US-10-332-929A-19
; Sequence 19, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-332-929A-19

Alignment Scores:

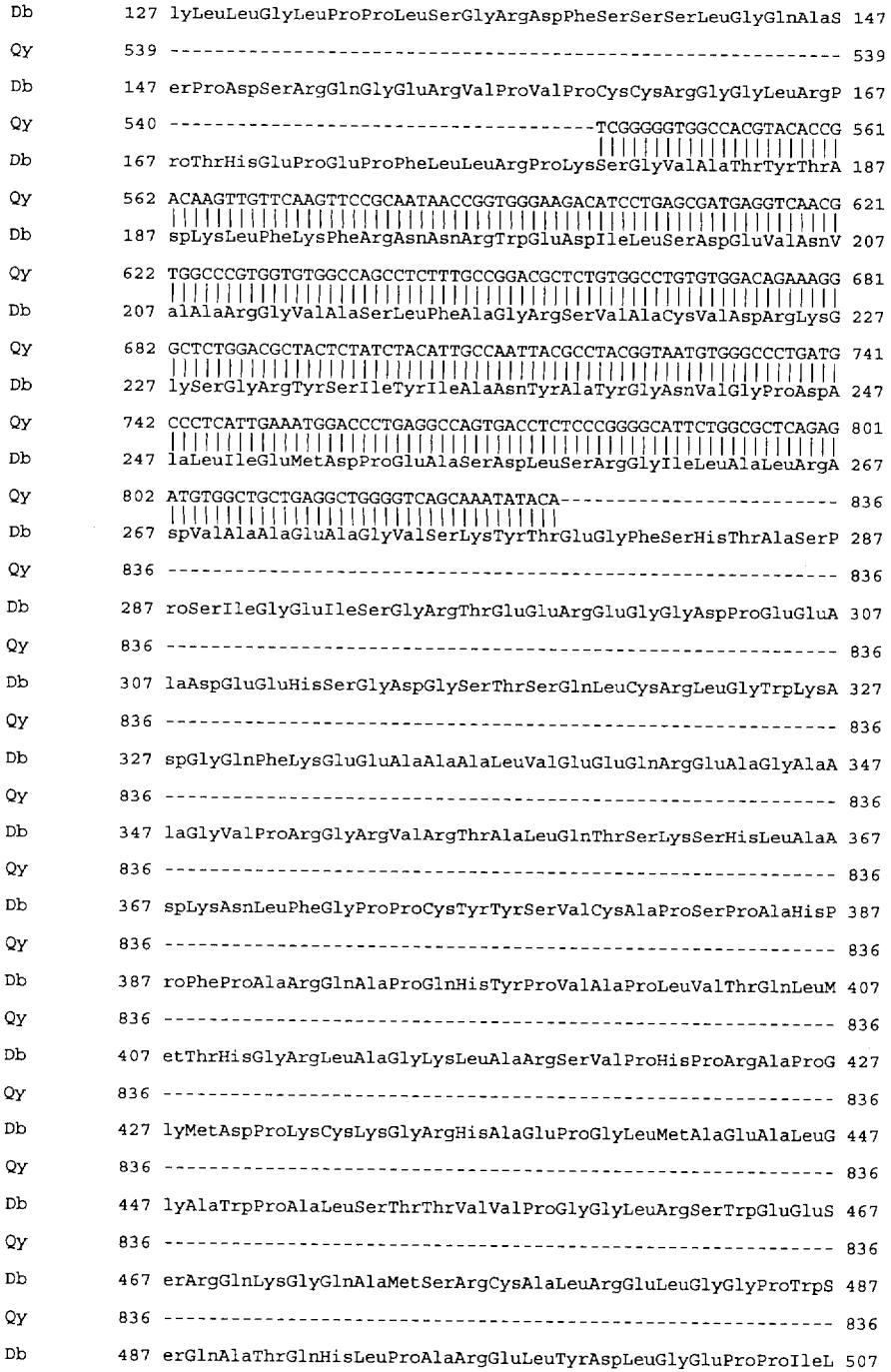
Pred. No.:	1.7e-201	Length:	637
Score:	3181.50	Matches:	602
Percent Similarity:	97.89%	Conservative:	2
Best Local Similarity:	97.57%	Mismatches:	6
Query Match:	78.44%	Indels:	7
DB:	12	Gaps:	1

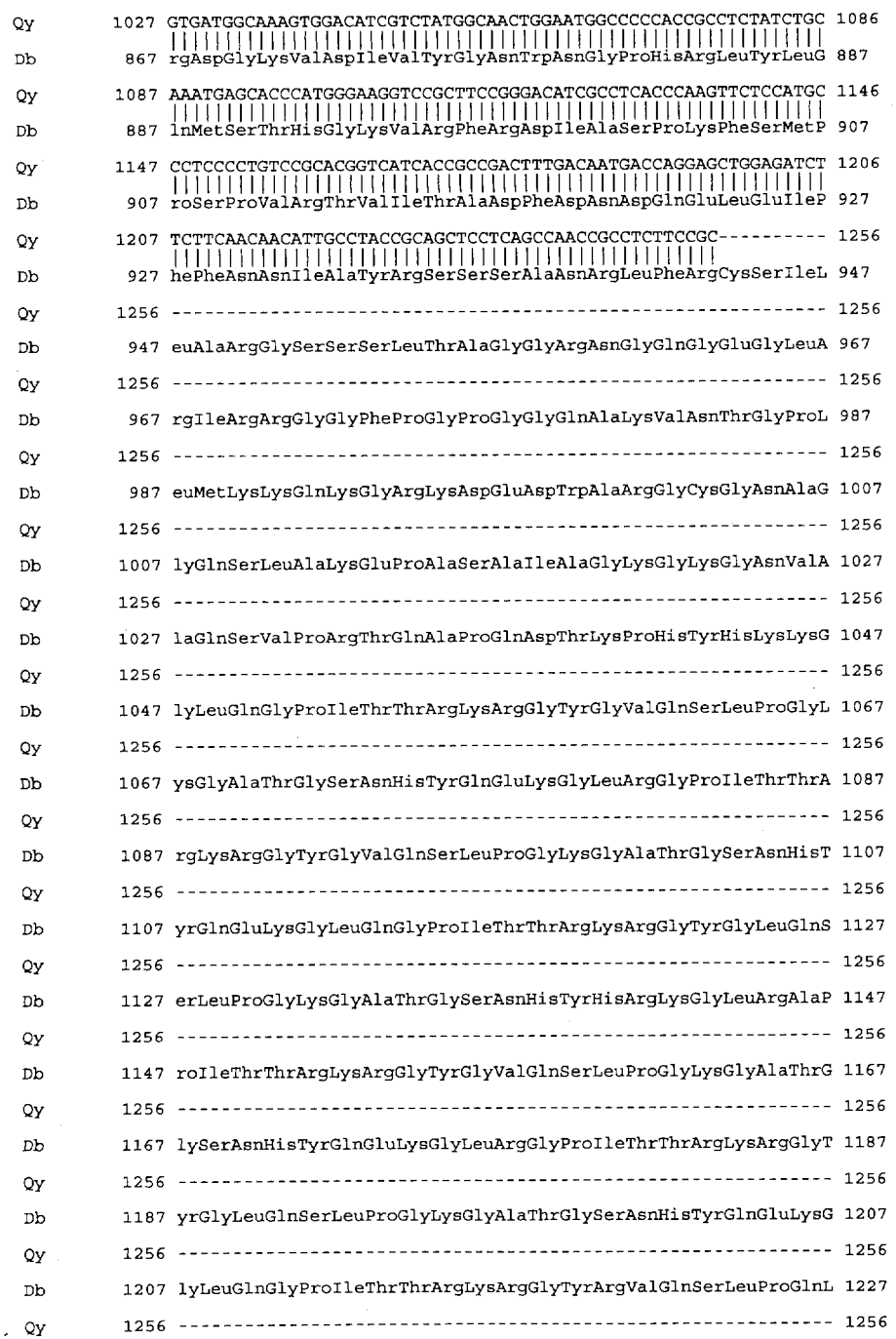
US-09-914-958B-35 (1-2177) x US-10-332-929A-19 (1-637)

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
```

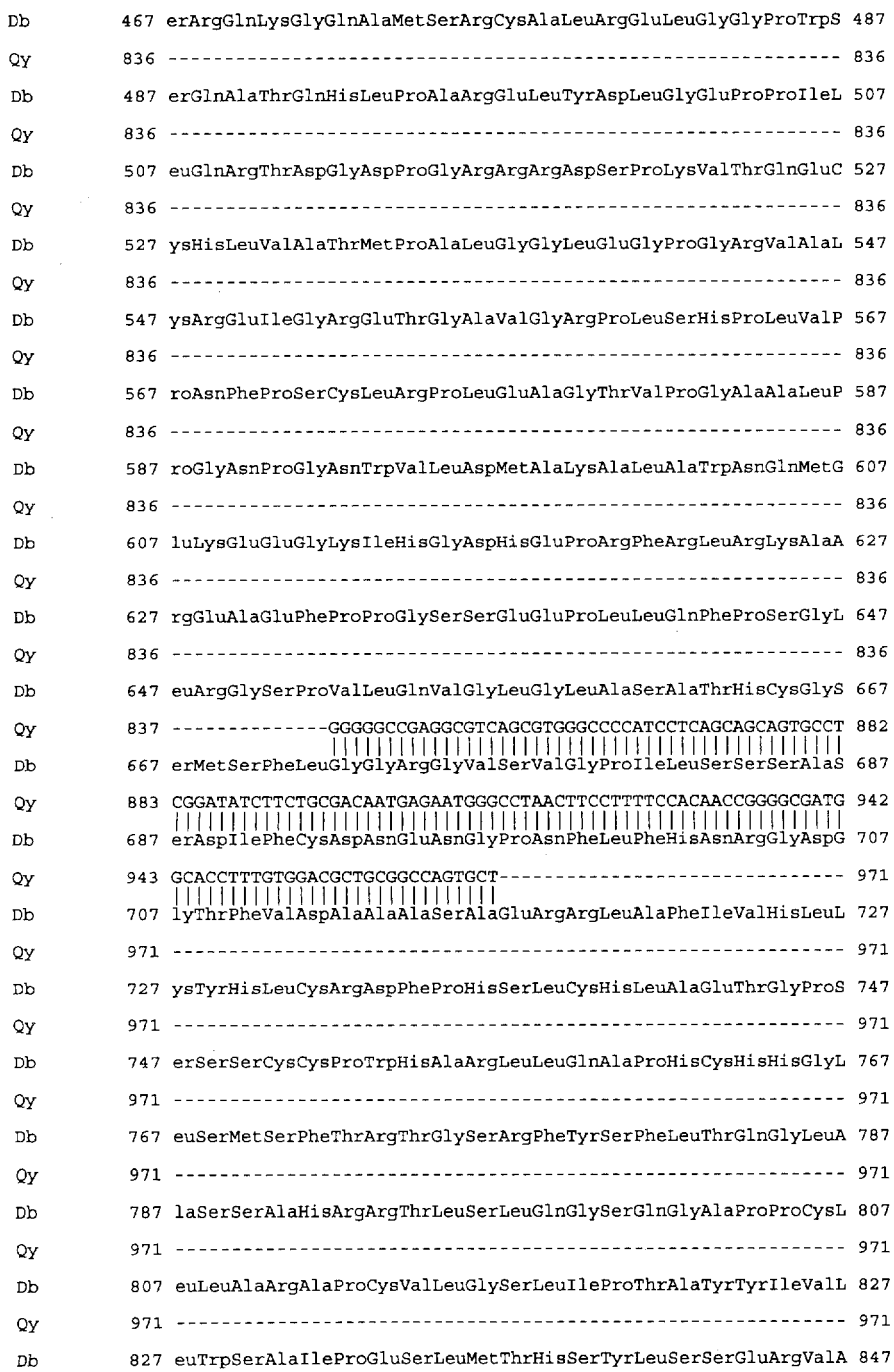
```
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346

Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406

Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466

Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526

Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586

Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646

Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706

Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766

Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826

Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886

Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG------ 1940

Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGln 600

Qy       1941 ----------------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976

Db        601 ValAlaPheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 617

RESULT 6
US-10-332-929A-21
; Sequence 21, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-332-929A-21

Alignment Scores:
Pred. No.:              2.46e-201       Length:         612
Score:                  3179.00         Matches:        598
Percent Similarity:     100.00%         Conservative:   0

Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            78.38%          Indels:         0
DB:                     12              Gaps:           0

US-09-914-958B-35 (1-2177) x US-10-332-929A-21 (1-612)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206

Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266

Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326

Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386

Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446

Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506

Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566

Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626

Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686

Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746

Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806

Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866

Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926

Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986

Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046

Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106

Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166

Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340
```

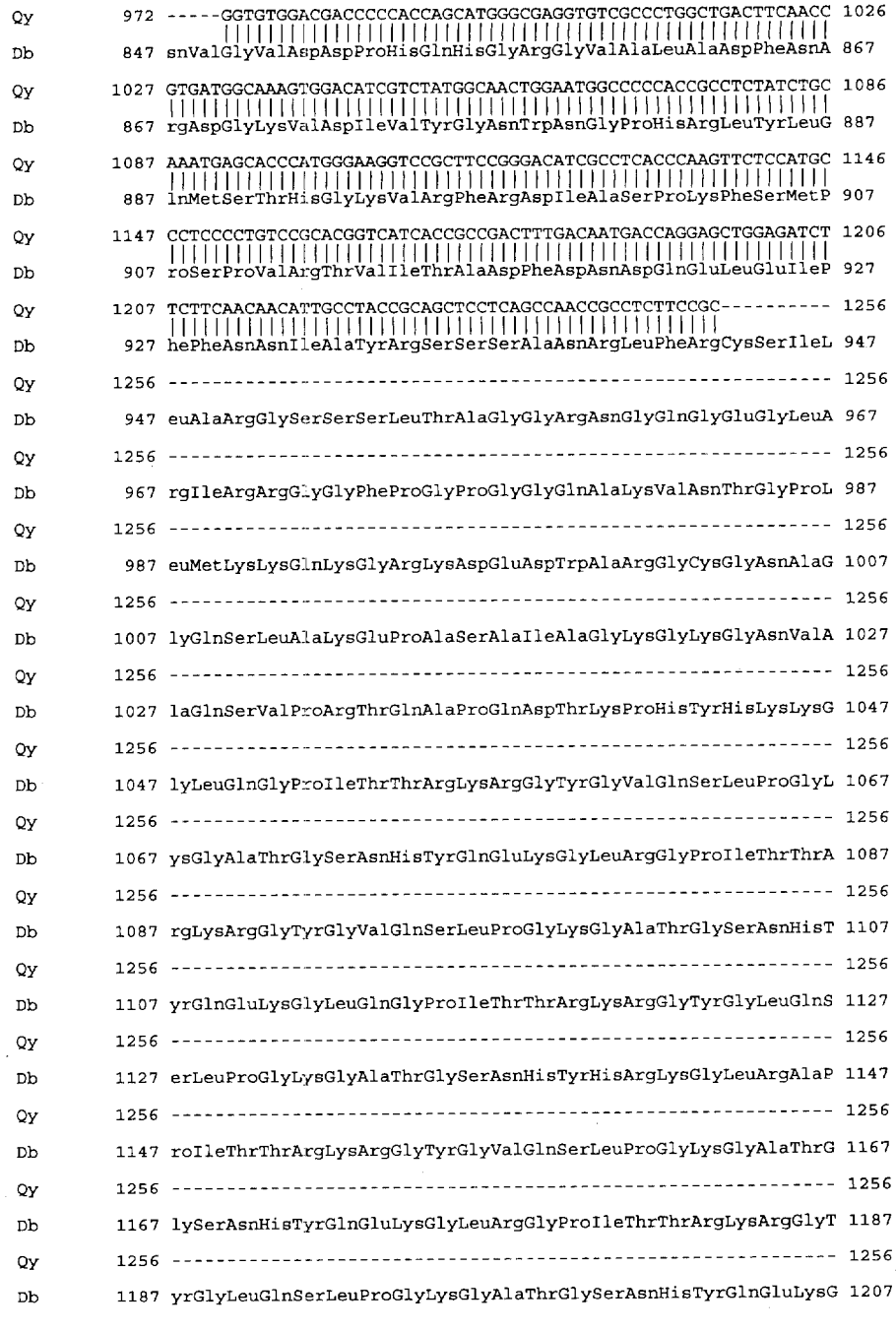
```
Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG 1940
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysVal 598

RESULT 7
US-10-295-027-185
; Sequence 185, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 185
;   LENGTH: 1597
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-185

Alignment Scores:
Pred. No.:                1.16e-145     Length:          1597
Score:                    2337.00       Matches:         566
Percent Similarity:       35.18%        Conservative:    6
Best Local Similarity:    34.81%        Mismatches:      29
Query Match:              57.62%        Indels:          1025
DB:                       15            Gaps:            11

US-09-914-958B-35 (1-2177) x US-10-295-027-185 (1-1597)

Qy        145 GCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGT 204
Db          2 AlaCysProGlyGlyLeu---ProAlaArgCysSerGlyTrpMet---GlyLeuGlyGly 19

Qy        205 CCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATG 264
Db         20 ProSerGlySerSerProAlaSer---------------------------------- 27

Qy        265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
Db         28 ---------ProPro---------------------------------------------- 29

Qy        325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
Db         30 ------HisSerSerSerArg--TyrAsnGlyProAsnLeuValLeuLysTyrAspArgA 47

Qy        385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
Db         47 laGlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuA 67

Qy        445 GGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
Db         67 rgAspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgG 87

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTC---------------------------- 539
Db         87 luGluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyHisSerSerSerAlaGlnV 107

Qy        539 ------------------------------------------------------------ 539
Db        107 alProSerGlyLeuHisArgAsnArgProValLeuLysProProProThrThrProAlaG 127

Qy        539 ------------------------------------------------------------ 539
```


Db 127 lyLeuLeuGlyLeuProProLeuSerGlyArgAspPheSerSerSerLeuGlyGlnAlaS 147
Qy 539 ---------- 539
Db 147 erProAspSerArgGlnGlyGluArgValProValProCysCysArgGlyGlyLeuArgP 167
Qy 540 ----------TCGGGGGTGGCCACGTACACCG 561
Db 167 roThrHisGluProGluProPheLeuLeuArgProLysSerGlyValAlaThrTyrThrA 187
Qy 562 ACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACG 621
Db 187 spLysLeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnV 207
Qy 622 TGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGG 681
Db 207 alAlaArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysG 227
Qy 682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATG 741
Db 227 lySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspA 247
Qy 742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
Db 247 laLeuIleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgA 267
Qy 802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACA---------- 836
Db 267 spValAlaAlaGluAlaGlyValSerLysTyrThrGluGlyPheSerHisThrAlaSerP 287
Qy 836 ---------- 836
Db 287 roSerIleGlyGluIleSerGlyArgThrGluGluArgGluGlyGlyAspProGluGluA 307
Qy 836 ---------- 836
Db 307 laAspGluGluHisSerGlyAspGlySerThrSerGlnLeuCysArgLeuGlyTrpLysA 327
Qy 836 ---------- 836
Db 327 spGlyGlnPheLysGluGluAlaAlaAlaLeuValGluGluGlnArgGluAlaGlyAlaA 347
Qy 836 ---------- 836
Db 347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAlaA 367
Qy 836 ---------- 836
Db 367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387
Qy 836 ---------- 836
Db 387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrGlnLeuM 407
Qy 836 ---------- 836
Db 407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427
Qy 836 ---------- 836
Db 427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447
Qy 836 ---------- 836
Db 447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467
Qy 836 ---------- 836
Db 467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487
Qy 836 ---------- 836
Db 487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProProIleL 507
Qy 836 ---------- 836
Db 507 euGlnArgThrAspGlyAspProGlyArgArgArgAspSerProLysValThrGlnGluC 527
Qy 836 ---------- 836
Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547
Qy 836 ---------- 836
Db 547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567
Qy 836 ---------- 836
Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587
Qy 836 ---------- 836
Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607
Qy 836 ---------- 836
Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627
Qy 836 ---------- 836
Db 627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647
Qy 836 ---------- 836
Db 647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlyS 667
Qy 837 ----------GGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS 687
Qy 883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707
Qy 943 GCACCTTTGTGGACGCTGCGGCCAGTGCT---------- 971
Db 707 lyThrPheValAspAlaAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727
Qy 971 ---------- 971
Db 727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747
Qy 971 ---------- 971
Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisHisGlyL 767
Qy 971 ---------- 971
Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787
Qy 971 ---------- 971
Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807
Qy 971 ---------- 971
Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827
Qy 971 ---------- 971
Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847
Qy 972 -----GGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACC 1026
Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867


Qy 1027 GTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGC 1086
Db 867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887
Qy 1087 AAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGC 1146
Db 887 lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP 907
Qy 1147 CCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCT 1206
Db 907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927
Qy 1207 TCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGC---------- 1256
Db 927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947
Qy 1256 ---------- 1256
Db 947 euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967
Qy 1256 ---------- 1256
Db 967 rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL 987
Qy 1256 ---------- 1256
Db 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007
Qy 1256 ---------- 1256
Db 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027
Qy 1256 ---------- 1256
Db 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047
Qy 1256 ---------- 1256
Db 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067
Qy 1256 ---------- 1256
Db 1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087
Qy 1256 ---------- 1256
Db 1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107
Qy 1256 ---------- 1256
Db 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127
Qy 1256 ---------- 1256
Db 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147
Qy 1256 ---------- 1256
Db 1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167
Qy 1256 ---------- 1256
Db 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187
Qy 1256 ---------- 1256
Db 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207
Qy 1256 ---------- 1256
Db 1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227
Qy 1256 ---------- 1256
Db 1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247
Qy 1256 ---------- 1256
Db 1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267
Qy 1256 ---------- 1256
Db 1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287
Qy 1257 ----------GTCATCCGTAGAGAGCACGGAGACCCCC 1284
Db 1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307
Qy 1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
Db 1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327
Qy 1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
Db 1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347
Qy 1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
Db 1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367
Qy 1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
Db 1367 alValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrL 1387
Qy 1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
Db 1387 ysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetG 1407
Qy 1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
Db 1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427
Qy 1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
Db 1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447
Qy 1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
Db 1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467
Qy 1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
Db 1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487
Qy 1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
Db 1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507
Qy 1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGC- 1943
Db 1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527
Qy 1944 ----------TGGTGGAGCCCTGTGTTGAAG---------- 1964
Db 1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547
Qy 1965 ----------ATAGTGACACCACAAG---------- 1980
Db 1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566
Qy 1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
Db 1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586
Qy 2041 CCTGTTGATCAG 2052

```
Db       1587 ProGluAlaGln 1590

RESULT 8
US-10-188-832-60
; Sequence 60, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
;  APPLICANT: Mack, David H.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
;  TITLE OF INVENTION:  and Methods of Screening for Modulators of Bladder
;  TITLE OF INVENTION:  Cancer
;  FILE REFERENCE: 018501-002330US
;  CURRENT APPLICATION NUMBER: US/10/188,832
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US 60/302,814
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: US 60/310,099
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/343,705
;  PRIOR FILING DATE: 2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/372,246
;  PRIOR FILING DATE: 2002-04-12
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
;   LENGTH: 1597
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-188-832-60

Alignment Scores:
Pred. No.:                1.16e-145      Length:          1597
Score:                    2337.00        Matches:         566
Percent Similarity:       35.18%         Conservative:    6
Best Local Similarity:    34.81%         Mismatches:      29
Query Match:              57.62%         Indels:          1025
DB:                       16             Gaps:            11

US-09-914-958B-35 (1-2177) x US-10-188-832-60 (1-1597)

Qy        145 GCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGT 204
              ||||||||||||                    |||||||||:::          |||   |||
Db          2 AlaCysProGlyGlyLeu---ProAlaArgCysSerGlyTrpMet---GlyLeuGlyGly 19

Qy        205 CCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATG 264
              |||||||||   :::|||   |||
Db         20 ProSerGlySerSerProAlaSer---------------------------------- 27

Qy        265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
                      ||||||
Db         28 ---------ProPro---------------------------------------------- 29

Qy        325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
                    ||||||   |||  ||||||||||||||||||||||||||||||||||
Db         30 ------HisSerSerSerArg--TyrAsnGlyProAsnLeuValLeuLysTyrAspArgA 47

Qy        385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 laGlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuA 67

Qy        445 GGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 rgAspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgG 87

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTC---------------------------- 539
              ||||||||||||||||||||||||||||||||
Db         87 luGluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyHisSerSerSerAlaGlnV 107

Qy        539 ------------------------------------------------------------ 539

Db        107 alProSerGlyLeuHisArgAsnArgProValLeuLysProProProThrThrProAlaG 127

Qy        539 ------------------------------------------------------------ 539

Db        127 lyLeuLeuGlyLeuProProLeuSerGlyArgAspPheSerSerSerLeuGlyGlnAlaS 147

Qy        539 ------------------------------------------------------------ 539

Db        147 erProAspSerArgGlnGlyGluArgValProValProCysCysArgGlyGlyLeuArgP 167

Qy        540 ------------------------------------TCGGGGGTGGCCACGTACACCG 561
                                                  ||||||||||||||||||||||
Db        167 roThrHisGluProGluProPheLeuLeuArgProLysSerGlyValAlaThrTyrThrA 187

Qy        562 ACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACG 621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 spLysLeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnV 207

Qy        622 TGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGG 681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 alAlaArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysG 227

Qy        682 GCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATG 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 lySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspA 247

Qy        742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 laLeuIleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgA 267

Qy        802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACA------------------------- 836
              ||||||||||||||||||||||||||||||||||
Db        267 spValAlaAlaGluAlaGlyValSerLysTyrThrGluGlyPheSerHisThrAlaSerP 287

Qy        836 ------------------------------------------------------------ 836

Db        287 roSerIleGlyGluIleSerGlyArgThrGluGluArgGluGlyGlyAspProGluGluA 307

Qy        836 ------------------------------------------------------------ 836

Db        307 laAspGluGluHisSerGlyAspGlySerThrSerGlnLeuCysArgLeuGlyTrpLysA 327

Qy        836 ------------------------------------------------------------ 836

Db        327 spGlyGlnPheLysGluGluAlaAlaAlaLeuValGluGluGlnArgGluAlaGlyAlaA 347

Qy        836 ------------------------------------------------------------ 836

Db        347 laGlyValProArgGlyArgValArgThrAlaLeuGlnThrSerLysSerHisLeuAlaA 367

Qy        836 ------------------------------------------------------------ 836

Db        367 spLysAsnLeuPheGlyProProCysTyrTyrSerValCysAlaProSerProAlaHisP 387

Qy        836 ------------------------------------------------------------ 836

Db        387 roPheProAlaArgGlnAlaProGlnHisTyrProValAlaProLeuValThrGlnLeuM 407

Qy        836 ------------------------------------------------------------ 836

Db        407 etThrHisGlyArgLeuAlaGlyLysLeuAlaArgSerValProHisProArgAlaProG 427

Qy        836 ------------------------------------------------------------ 836

Db        427 lyMetAspProLysCysLysGlyArgHisAlaGluProGlyLeuMetAlaGluAlaLeuG 447

Qy        836 ------------------------------------------------------------ 836

Db        447 lyAlaTrpProAlaLeuSerThrThrValValProGlyGlyLeuArgSerTrpGluGluS 467

Qy        836 ------------------------------------------------------------ 836
```


Db 467 erArgGlnLysGlyGlnAlaMetSerArgCysAlaLeuArgGluLeuGlyGlyProTrpS 487
Qy 836 ---------- 836
Db 487 erGlnAlaThrGlnHisLeuProAlaArgGluLeuTyrAspLeuGlyGluProProIleL 507
Qy 836 ---------- 836
Db 507 euGlnArgThrAspGlyAspProGlyArgArgArgAspSerProLysValThrGlnGluC 527
Qy 836 ---------- 836
Db 527 ysHisLeuValAlaThrMetProAlaLeuGlyGlyLeuGluGlyProGlyArgValAlaL 547
Qy 836 ---------- 836
Db 547 ysArgGluIleGlyArgGluThrGlyAlaValGlyArgProLeuSerHisProLeuValP 567
Qy 836 ---------- 836
Db 567 roAsnPheProSerCysLeuArgProLeuGluAlaGlyThrValProGlyAlaAlaLeuP 587
Qy 836 ---------- 836
Db 587 roGlyAsnProGlyAsnTrpValLeuAspMetAlaLysAlaLeuAlaTrpAsnGlnMetG 607
Qy 836 ---------- 836
Db 607 luLysGluGluGlyLysIleHisGlyAspHisGluProArgPheArgLeuArgLysAlaA 627
Qy 836 ---------- 836
Db 627 rgGluAlaGluPheProProGlySerSerGluGluProLeuLeuGlnPheProSerGlyL 647
Qy 836 ---------- 836
Db 647 euArgGlySerProValLeuGlnValGlyLeuGlyLeuAlaSerAlaThrHisCysGlyS 667
Qy 837 --------------GGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
Db 667 erMetSerPheLeuGlyGlyArgGlyValSerValGlyProIleLeuSerSerSerAlaS 687
Qy 883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
Db 687 erAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspG 707
Qy 943 GCACCTTTGTGGACGCTGCGGCCAGTGCT---------- 971
Db 707 lyThrPheValAspAlaAlaAlaSerAlaGluArgArgLeuAlaPheIleValHisLeuL 727
Qy 971 ---------- 971
Db 727 ysTyrHisLeuCysArgAspPheProHisSerLeuCysHisLeuAlaGluThrGlyProS 747
Qy 971 ---------- 971
Db 747 erSerSerCysCysProTrpHisAlaArgLeuLeuGlnAlaProHisCysHisHisGlyL 767
Qy 971 ---------- 971
Db 767 euSerMetSerPheThrArgThrGlySerArgPheTyrSerPheLeuThrGlnGlyLeuA 787
Qy 971 ---------- 971
Db 787 laSerSerAlaHisArgArgThrLeuSerLeuGlnGlySerGlnGlyAlaProProCysL 807
Qy 971 ---------- 971
Db 807 euLeuAlaArgAlaProCysValLeuGlySerLeuIleProThrAlaTyrTyrIleValL 827
Qy 971 ---------- 971
Db 827 euTrpSerAlaIleProGluSerLeuMetThrHisSerTyrLeuSerSerGluArgValA 847


Qy 972 -----GGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACC 1026
Db 847 snValGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnA 867
Qy 1027 GTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGC 1086
Db 867 rgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuG 887
Qy 1087 AAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGC 1146
Db 887 lnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSerMetP 907
Qy 1147 CCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCT 1206
Db 907 roSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGluIleP 927
Qy 1207 TCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGC---------- 1256
Db 927 hePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgCysSerIleL 947
Qy 1256 ---------- 1256
Db 947 euAlaArgGlySerSerSerLeuThrAlaGlyGlyArgAsnGlyGlnGlyGluGlyLeuA 967
Qy 1256 ---------- 1256
Db 967 rgIleArgArgGlyGlyPheProGlyProGlyGlyGlnAlaLysValAsnThrGlyProL 987
Qy 1256 ---------- 1256
Db 987 euMetLysLysGlnLysGlyArgLysAspGluAspTrpAlaArgGlyCysGlyAsnAlaG 1007
Qy 1256 ---------- 1256
Db 1007 lyGlnSerLeuAlaLysGluProAlaSerAlaIleAlaGlyLysGlyLysGlyAsnValA 1027
Qy 1256 ---------- 1256
Db 1027 laGlnSerValProArgThrGlnAlaProGlnAspThrLysProHisTyrHisLysLysG 1047
Qy 1256 ---------- 1256
Db 1047 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyL 1067
Qy 1256 ---------- 1256
Db 1067 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1087
Qy 1256 ---------- 1256
Db 1087 rgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisT 1107
Qy 1256 ---------- 1256
Db 1107 yrGlnGluLysGlyLeuGlnGlyProIleThrThrArgLysArgGlyTyrGlyLeuGlnS 1127
Qy 1256 ---------- 1256
Db 1127 erLeuProGlyLysGlyAlaThrGlySerAsnHisTyrHisArgLysGlyLeuArgAlaP 1147
Qy 1256 ---------- 1256
Db 1147 roIleThrThrArgLysArgGlyTyrGlyValGlnSerLeuProGlyLysGlyAlaThrG 1167
Qy 1256 ---------- 1256
Db 1167 lySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrArgLysArgGlyT 1187
Qy 1256 ---------- 1256
Db 1187 yrGlyLeuGlnSerLeuProGlyLysGlyAlaThrGlySerAsnHisTyrGlnGluLysG 1207

```
Qy       1256 ------------------------------------------------------------ 1256
Db       1207 lyLeuGlnGlyProIleThrThrArgLysArgGlyTyrArgValGlnSerLeuProGlnL 1227
Qy       1256 ------------------------------------------------------------ 1256
Db       1227 ysGlyAlaThrGlySerAsnHisTyrGlnGluLysGlyLeuArgGlyProIleThrThrA 1247
Qy       1256 ------------------------------------------------------------ 1256
Db       1247 rgLysArgGlyTyrGlyLeuGlnSerLeuProGlyLysGluAlaMetGlySerAsnHisT 1267
Qy       1256 ------------------------------------------------------------ 1256
Db       1267 yrGlnGluLysGlyLeuArgAlaProIleThrThrArgLysArgGlyTyrGlyValGlnS 1287
Qy       1257 ------------------------------GTCATCCGTAGAGAGCACGGAGACCCCC 1284
                                              ||||||||||||||||||||||||||||
Db       1287 erLeuProGlnLysGlyAlaThrGlySerAsnValIleArgArgGluHisGlyAspProL 1307
Qy       1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1307 euIleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyV 1327
Qy       1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 alValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerM 1347
Qy       1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1347 etAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgV 1367
Qy       1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1367 alValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrL 1387
Qy       1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 ysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetG 1407
Qy       1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1407 luProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpP 1427
Qy       1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1427 roAspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluI 1447
Qy       1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 leLeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnG 1467
Qy       1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1467 lyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProP 1487
Qy       1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1487 heValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgT 1507
Qy       1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGC- 1943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db       1507 hrAsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGlyT 1527
Qy       1944 -------------------------TGGTGGAGCCCTGTGTTGAAG------------- 1964
                                         |||   |||     |||
Db       1527 hrGluLeuGlySerArgHisThrMetThrTrpLysProArgProLysLysGluLeuGlnL 1547
Qy       1965 -----------ATAGTGACACCACAAG------------------------------- 1980
                       |||   ||||||
Db       1547 euSerGlnGlyIleCysThrPro-ValTrpSerPhePheLeuProGlyCysArgLeuLeu 1566
Qy       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
                  |||   ||||||   :::     :::   :::     |||:::     |||||||||
Db       1567 LeuLysArgAlaGlnLeuGlnAlaAlaProSerThrLeuLeuGlnLysAlaProGlyIle 1586

Qy       2041 CCTGTTGATCAG 2052
              |||     |||
Db       1587 ProGluAlaGln 1590
```

RESULT 9
US-10-295-027-181
; Sequence 181, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 181
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-181

Alignment Scores:

Pred. No.:	2.1e-121	Length:	418
Score:	1967.50	Matches:	385
Percent Similarity:	89.81%	Conservative:	3
Best Local Similarity:	89.12%	Mismatches:	10
Query Match:	48.51%	Indels:	34
DB:	15	Gaps:	2

US-09-914-958B-35 (1-2177) x US-10-295-027-181 (1-418)

```
Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40
```

```
Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu----------------- 334

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 -------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
                                                                  |||   ::
Db        384 ----------------------------------------------------LeuSerAlaSe 387

Qy       2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
              :|||        :::||||||      |||:::
Db        387 rHisArgAlaProAlaProProProProProLeu 398
```

RESULT 10
US-10-188-832-56
; Sequence 56, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-188-832-56

Alignment Scores:

Pred. No.:	2.1e-121	Length:	418
Score:	1967.50	Matches:	385
Percent Similarity:	89.81%	Conservative:	3
Best Local Similarity:	89.12%	Mismatches:	10
Query Match:	48.51%	Indels:	34
DB:	16	Gaps:	2

US-09-914-958B-35 (1-2177) x US-10-188-832-56 (1-418)

```
Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120
```

```
Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
             ||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
             ||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu---------------- 334

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
                                      ||||||||||||||||||||||||||||||||||
Db        335 ------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
                                                                |||   ::
Db        384 ----------------------------------------------------LeuSerAlaSe 387

Qy       2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
              :|||         :::|||||||      |||:::
Db        387 rHisArgAlaProAlaProProProProProLeu 398
```

RESULT 11
US-09-892-877-247
; Sequence 247, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: PZ028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-892-877-247

Alignment Scores:

Pred. No.:	4.77e-90	Length:	306
Score:	1492.00	Matches:	291
Percent Similarity:	96.38%	Conservative:	2
Best Local Similarity:	95.72%	Mismatches:	7
Query Match:	36.79%	Indels:	7
DB:	10	Gaps:	2

US-09-914-958B-35 (1-2177) x US-09-892-877-247 (1-306)

```
Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db          1 TrpGlnSerGlyHisArgLeuTrpGlnLeuGluTrpProProProProLeuSerAlaAsp 20

Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC---TCACCCAAGTTCTCCATGCC 1147
             |||||||||||||||||||||||||||||||    :::   ||||||||||||||||||||
Db         21 GluHisProTrpGluGlyProLeuProGly-ThrSerProSerProLysPheSerMetPr 40

Qy       1148 CTCCCCTGT-CCGCACGGTCATCACCG-CCGACTTTGACAATGACCAGGAGCTGGAGATC 1205
             |||||||   |||||||||||||||   ||||||||||||||||||||||||||||||||
Db         40 oSerProValProHisGlyHisHisArgProThrLeuThrMetThrArgSerTrpArgIl 60

Qy       1206 -TTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ePhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgValIleAr 80

Qy       1265 TAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGA 1324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 gArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGluProGl 100

Qy       1325 GGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCAT 1384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 uGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAspLeuIl 120

Qy       1385 CTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTT 1444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 eLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPh 140

Qy       1445 CAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGC 1504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 eAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArgGlyAl 160

Qy       1505 TAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTC 1564
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 aLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySe 180

Qy       1565 AGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAG 1624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 rGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAlaSe 200

Qy       1625 CAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGA 1684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 rSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGlyGl 220

Qy       1685 GATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGC 1744
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 uMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspProAl 240
```

```
Qy       1745 CCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAA 1804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 aProLeuGluCysGlyGlnGlyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAs 260

Qy       1805 TGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTA 1864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 nGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsnThrTy 280

Qy       1865 TGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAG---CCCAACGA 1921
             |||||||||||||||||||||||||||||||||||   |||         ||| ||||
Db        280 rGlySerTyrArgCysArgThrAsnLysLysCysSer***GlyLeuArgValPro-ThrA 300

Qy       1922 GGATGGCA 1929
             ||||||||
Db        300 rgMetAla 302
```

RESULT 12
US-09-948-783-249
; Sequence 249, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: PZ028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 249
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-948-783-249

Alignment Scores:

Pred. No.:	4.77e-90	Length:	306
Score:	1492.00	Matches:	291
Percent Similarity:	96.38%	Conservative:	2
Best Local Similarity:	95.72%	Mismatches:	7
Query Match:	36.79%	Indels:	7
DB:	10	Gaps:	2

US-09-914-958B-35 (1-2177) x US-09-948-783-249 (1-306)

```
Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db          1 TrpGlnSerGlyHisArgLeuTrpGlnLeuGluTrpProProProProLeuSerAlaAsp 20

Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC---TCACCCAAGTTCTCCATGCC 1147
             |||||||||||||||||||||||||||||||    :::   ||||||||||||||||||||
Db         21 GluHisProTrpGluGlyProLeuProGly-ThrSerProSerProLysPheSerMetPr 40

Qy       1148 CTCCCCTGT-CCGCACGGTCATCACCG-CCGACTTTGACAATGACCAGGAGCTGGAGATC 1205
             |||||||   |||||||||||||||   |||||||||||||||||||||||||||||||||
Db         40 oSerProValProHisGlyHisHisArgProThrLeuThrMetThrArgSerTrpArgIl 60

Qy       1206 -TTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCG 1264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 ePhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgValIleAr 80

Qy       1265 TAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGA 1324
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 gArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGluProGl 100

Qy       1325 GGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCAT 1384
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 uGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAspLeuIl 120

Qy       1385 CTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTT 1444
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 eLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGlyPh 140

Qy       1445 CAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGC 1504
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 eAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArgGlyAl 160

Qy       1505 TAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTC 1564
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 aLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySe 180

Qy       1565 AGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAG 1624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 rGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAlaSe 200

Qy       1625 CAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGA 1684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 rSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGlyGl 220

Qy       1685 GATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGC 1744
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 uMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspProAl 240

Qy       1745 CCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAA 1804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 aProLeuGluCysGlyGlnGlyPheSerGlnGlnGluAsnGlyHisCysMetAspThrAs 260

Qy       1805 TGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTA 1864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 nGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsnThrTy 280

Qy       1865 TGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAG---CCCAACGA 1921
             |||||||||||||||||||||||||||||||||||   |||         ||| ||||
Db        280 rGlySerTyrArgCysArgThrAsnLysLysCysSer***GlyLeuArgValPro-ThrA 300

Qy       1922 GGATGGCA 1929
             ||||||||
Db        300 rgMetAla 302
```

RESULT 13
US-10-084-846A-6
; Sequence 6, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 6
; LENGTH: 19662
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6

Alignment Scores:

Pred. No.:	3.83e-12	Length:	19662
Score:	318.00	Matches:	189
Percent Similarity:	32.75%	Conservative:	36
Best Local Similarity:	27.51%	Mismatches:	232
Query Match:	7.69%	Indels:	230
DB:	15	Gaps:	44

US-09-914-958B-35 (1-2177) x US-10-084-846A-6 (1-19662)

```
Qy       1833 GGCACACGA---ATGGGAACTGGATGCATTCATTGGTGTCCATGCAATGGCCATTTTCCT 1777
Db        607 GlyThrArgProIleGlyAlaAlaCys-------------------------------- 615

Qy       1776 GCTGGGAGAATCCTTGGCCACACTCCA-----------GTGGGGCTGGGTCCTGAAGTG 1729
Db        616 ---------ValAlaGlyAspThrProCysSerLysGlyLeuGlyGlyGlyThrProTyr 632

Qy       1728 TGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGGCCACGT 1669
Db        633 ProProArgProAlaGlyLeuAspSerLeuAlaGlyGlyGlyValPro---GlyProAsp 651

Qy       1668 TCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGC--------------- 1624
Db        652 ValGlyArgProArgGlyTrpProGlyArgAspProValProCysArgGluAspAspAsp 671

Qy       1623 TGGCTTCATCCTTCCCCAGGCCAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGCCTG 1564
Db        672 Trp------------SerGlyArgAlaArgThrAlaAlaProArgArg---------Leu 684

Qy       1563 AGCCCCCGTCGATGATCCTCA------GGTGGGCCCCACTCTTCTTGGTGTAGAGCACGA 1510
Db        685 CysAlaArgCysArgSerSerArgCysArgSerProAlaValSerThrCysArgTrpPro 704

Qy       1509 CCTTAGCTCCCCTGGCAAAGG---CCCCAAACCGGGTGCGTGGCACCACTCGCAGCCAGT 1453
Db        705 Pro------ProArgArgArgSerProSerThrTrpCys--------------------- 715

Qy       1452 TGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCAT 1393
Db        716 CysCys------ProAsnCysArgArgCysArgThrThrAlaCysThrProCysArgArg 733

Qy       1392 GGGACAAGATGAGGTCCAGCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCCCTGTGC 1333
```

```
Db        734 GlyArgSerAlaGlyProGlyGlyArgArgCysThrCysGlyArgCysGlyProLeuArg 753

Qy       1332 CCCGGCCCTCAGGCTCCA---AGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTC 1276
Db        754 ArgGlyProArgAlaArgCysArgArgArgArgArgProSerPro---------------- 768

Qy       1275 CGTGCTCTCTACGGATGACGCGGAAGAGGCGGTTGG---CTGAGGAGCTGCGGTAGGCAA 1219
Db        769 --------------------GlyThrArgProTrpSerCysArgSerCysSerHisArg 781

Qy       1218 TGTTGTTGAAGAAGATCTCCAGCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGC 1159
Db        782 CysSerArgSerAlaGlyProGluGlyGlyArgAlaAlaGlyArgArg---------Thr 798

Qy       1158 GGACAGGGG---------------AGGGCATGG---AGAACTTGGGTGAGGCGATGT--- 1120
Db        799 GlyAlaGlyTrpAlaSerProCysArgGlyTrpArgArgArgTrpTrpArgArgThrSer 818

Qy       1119 ------------CCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGCAGATAGAGGCGGT 1072
Db        819 AlaProTrpArgProGlyThrGlyTrpProArgGlyTrpArg---------ArgGlyGly 835

Qy       1071 ----------------------GGGGGCCATTCCAGTTGCCATAGACGATGTCCACTT 1036
Db        836 AlaArgArgArgProProArgAlaGlySerProSerThrTyrArgArgTrpCysTrpSer 855

Qy       1035 TGCCATCACGGTTGAAGTCAGCCA----------------------GGGCGACACCTC 1000
Db        856 ProSerArgLeuSerSerArgProArgTrpThrAsnArgGlyArgArgGlyArgGlyArg 875

Qy        999 GCCCATGCTGGTGGGGGTCGTCCA--------------------CACCAGCACTGGCCG 961
Db        876 SerArgSerTrpSerSerArgProTrpArgProSerTrpSerSerArgArgProTrpPro 895

Qy        961 ------------------------------------------------------------ 961
Db        896 ArgArgValArgValThrSerTrpArgArgTrpSerThrArgCysArgProProArgArg 915

Qy        960 ------CAGCGTCCACAAAGGTGCCATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCAT 907
Db        916 ArgArgArgArgProTrpArgCysValArgProCysArgGlyPro---------------- 930

Qy        906 TCTCATTGTCGCAGAAGATATCCGAGGCACTGCTGCTGAGGATGG---GGCCCACGCTGA 850
Db        931 ------CysArgArgArgArgArgGlyThrCys---------TrpArgArgProAlaArg 945

Qy        849 CGCCTCGGCCCCCTGTATATTTGCTGACCCCAGCCTCAGCAGCCACATCTCTGAGCGCCA 790
Db        946 ArgSerGlyArgProSerArgCysArgArgArgProGlyArg------------------ 959

Qy        789 GAATGCCCCGGGAGAGGTCACTGGCCTCAGGGTCCATTTCAATGAGGGCATCAGGGCCCA 730
Db        960 TrpCysProGlyArgSerCysSerAlaArgGlyArgLeuArg---------SerTrpPro 976

Qy        729 CATTACCGTAGGCGTAATTGGCAATGTAGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCA 670
Db        977 SerTrpArgArgArgArg---ArgCysArg-----------------------ArgPro 987

Qy        669 CACAGGCCACAGAGCGTCCGGCAAAGAGGCTGGCCACACCACGGGCCACGTTGACCTCAT 610
Db        988 CysArgProArgAlaCysArgGlyArgGlyArgPro----------------------- 999

Qy        609 CGCTCAGGATGTCTTCCCACCGGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGG 550
Db       1000 --------------------GlyArgCys--------------CysArgCysArgTrp 1007

Qy        549 CCACCCCCGAGAAGGCATTATTGGTGTTGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGT 490
Db       1008 ArgArgSerArgArgArgPro---------GlyArgArgSerArgCysGlyCysGlyArg 1024

Qy        489 CGATGTCGCAGGCTGTGACCCCAATGGCGTTCCCCTGCCGGTCCCGCAGCGCGTAGTAGG 430
Db       1025 ArgCysArgArgThr---ProSer------LeuProAlaArgProSerProGlyArgCys 1041
```

```
Qy        429 GTGAGCTGCGCTCATCGACCGCGATGTTCACCAGCCGCTTCTGGGCCCGGTCATACTTCA 370
Db       1042 ThrSerCysSerGlyGlyGlySerSerGlyProArgGlyProGlyProArgArgAlaGly 1061

Qy        369 GAACCAGGTTGGGTCCATTGTACC----------------------CCGCCACGACGA 334
Db       1062 TrpSerGlySerArgArgCysThrCysGlySerGlyProArgArgArgProProGlyArg 1081

Qy        333 TCTCAAAGTCCCCATCATGGTCCACATCAGTAACTGCCACACCATAGTTGAGCTGGGTGG 274
Db       1082 ArgArgAlaPro---------------------------------------------- 1085

Qy        273 GATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTGAACATGGGTTCAG 214
Db       1086 ----------------------------------------------------TrpProPro 1088

Qy        213 CCCGCTGGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGCAGGA---ACGGTAACA 157
Db       1089 ProArgArg--------------ProArgSerArgSerArgSerGlyProSerCysThr 1103

Qy        156 TCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCGGCCCCGCCGC--- 100
Db       1104 CysAlaArgCysArg---ArgArgThrSerArgArgGlyArgValArgProArgAsnPro 1122

Qy         99 CTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCGCCTCTGCCGCCGGCGCCGCCGCCTGCTT 40
Db       1123 AsnGlyArgGlyProProGlySerArgCysSer---CysArgProArgArg--------- 1138

Qy         39 GCTCCCAGCCCCGGTCCCGGG 19
Db       1139 LeuProArgProArgThrGly 1145
```

RESULT 14
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:

Pred. No.:	1.29e-11	Length:	19723
Score:	310.00	Matches:	225
Percent Similarity:	29.33%	Conservative:	56
Best Local Similarity:	23.49%	Mismatches:	284
Query Match:	7.64%	Indels:	393
DB:	15	Gaps:	54

US-09-914-958B-35 (1-2177) x US-10-084-846A-5 (1-19723)

```
Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
Db       6423 ArgCysSerSerGlySerProArgCysGly---GlyHisGlyAlaSerSerProProPro 6441

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
Db       6442 SerSerAlaThrArgSerValArgSerProProProSerProProGly------------ 6457

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGA---TGTTACCGTTCCTGCTGCTGC 177
Db       6458 --------------ArgProProThrGlyProGlySerAlaValGlyArgCysCysCys 6472

Qy        178 TCT-----------------------------GGTTTCTGCCCATCACTGAGGGGTCCC 207
Db       6473 AlaArgProArgGlyValAlaArgTrpArgTrpSerValCysProSerThrArgProPro 6492

Qy        208 AGCGGGCTGAACCCATGTTCACTGCAGTCA------------CCAACTCAGTTCTGCCTC 255
Db       6493 SerGlySerArgAlaMetThrArgTrpSerArgProSerArgProArgArg--------- 6509

Qy        256 CTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACC 315
Db       6510 --------------ArgProProSerProValThrProAlaArgSerArgArgSerSer 6524

Qy        316 ATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGT 375
Db       6525 AlaAlaGlyProThrArgGlyTrpSerTyrGlyGlySerProProThrTrp--------- 6541

Qy        376 ATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACT 435
Db       6542 ------------------------------------ProSerThrAlaArgThrTrp 6548

Qy        436 ACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCG------------ 483
Db       6549 ThrProCysSerThrGlyCysAlaArgProProThrSerValProAlaHisArgThrArg 6568

Qy        484 ------------------------------------ACATCGACGGGGACGGCCGGG 504
Db       6569 ArgSerThrArgArgArgSerArgThrArgGlyProArgProProThrAlaArgThrGly 6588

Qy        505 AGGAGATCTACTTCC------------TCAACACCAATAATGCCTTCTCGGGGGTGGCCA 552
Db       6589 ArgArgThrCysAlaThrArgCysGlySerProProArgSerProArgArg------Pro 6606

Qy        553 CGTACACCGACAAGTTGTTCAAGTTCCGCA------ATAACCGGTGGGAAGACATCCTGA 606
Db       6607 ArgThrAlaThrGlyArgSerSerSerCysProArgThrProTrpSerArgThr------ 6624

Qy        607 GCGATGAGGTCAACG------TGGCCC---GTGGTGTGG---------------------- 636
Db       6625 ------ArgSerThrArgArgTrpProAsnAlaAlaTrpArgThrCysSerSerAspArg 6642

Qy        637 ------------------------------------CCAGCCTCTTTGCCGGACGCT 657
Db       6643 ArgTyrGlyAlaThrSerProArgProGlyProSerAlaProProTrpAlaProProThr 6662

Qy        658 CTGTGGCCTGTGTGGACA---------------------------------------- 675
Db       6663 AlaThrAlaAlaTrpThrGlyArgArgSerSerArgThrGlyThrSerLysSerCysArg 6682

Qy        676 ------------------------------------------------GAAAGGGCTCTG 687
Db       6683 ProThrProGlySerThrAlaArgCysGlyValProSerProGlyArgGluArgProSer 6702

Qy        688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG--------- 738
Db       6703 AlaAlaThrThrSer------------ThrProThrArgCysTrpAlaArgProAlaAla 6718

Qy        739 ---------------------ATGCCCTCATTGAAATGGACCCTGAGGCCA--------- 768
Db       6719 SerArgAlaAlaThrCysGlyCysGlyThrAlaArgTrpThrThrThrAlaAlaArgThr 6738

Qy        769 ---GTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG---------ATGTGGCTGCTGAGG 816
Db       6739 ArgAlaAlaThrProSerThrAlaTrpArgSerSerArgProProCysTrpArgSerArg 6758
```

```
Qy       817 CTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC--------- 867
Db      6759 ----------------------------------SerTrpArgProAlaProArgAla 6766

Qy       868 ---TCAGCAGCAGTGCCTCGGATA---------------------------------- 888
Db      6767 LysSerAlaAlaProSerArgThrArgProThrArgCysArgArgValSerGlyArgSer 6786

Qy       889 ---TCTTCTGCGACAATGAGAATGGGCCTAACTTCCT----------------------- 922
Db      6787 ArgSerSerAlaArgAlaArg-TrpCysGlyTrpProProGlyArgSerArgThrProPr 6806

Qy       923 -------------------------TTTCCACAACCGGGGCGA----------------- 940
Db      6806 oThrArgThrProProGlySerSerMetProArgProGlyArgProArgArgThrSerPr 6826

Qy       941 ----TGGCACCTTTGTGGACGCTGCGGCCAGTGCTGG------------------TGTGG 978
Db      6826 oValTrp------------ArgArgGlyArgCysTrpThrArgAlaSerThrGlySerAs 6842

Qy       979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
Db      6842 nArgProThrProAlaTrpSer---------ProGly---------------------- 6851

Qy      1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
Db      6852 ----------------------GlyTrpPro--------------------------Ar 6855

Qy      1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
Db      6855 gTrpAlaTyrProArgPro--ValSerIleGlyAlaSerSerAspCysProProValSer 6874

Qy      1159 GCA------------CGGTCATCACCGCCGACT---TTGACAATGACCAGGAGCTGGAGA 1203
Db      6875 ValTyrCysThrLeuArgCysSerArgProThrProArgProGlyProArgCysTrpThr 6894

Qy      1204 TCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCC 1263
Db      6895 ProCysArgSerArgArgProProSer------TrpAlaSerArgSerSerAlaTrpSer 6912

Qy      1264 GTAGAGAGCACGGAGACCCCCTCA-------------------------TCGAGGAGC 1296
Db      6913 CysThrSerThrArgSerProSerThrAlaArgHisArgArgArgArgArgSerArgSer 6932

Qy      1297 ---TCAATCCCGGCG---------------ACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
Db      6933 ArgSerIleProAlaSerProThrProCysThrProTrpSer---ArgThrGlyArgAsp 6951

Qy      1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
Db      6952 AlaArgTrpArgAlaCysAlaAlaCysAlaThrArgTrpSerSerSerArgProArgArg 6971

Qy      1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
Db      6972 ThrPro--------------------------ThrSerArgAlaAlaThrArgThrArg 6982

Qy      1459 TGCGAGTGGTGCCACGCA--------------------------CCCGGTTTGGGGCCT 1491
Db      6983 ---ThrTrpCysArgSerArgValCysArgArgArgSerCysValProGlyCysSerThr 7001

Qy      1492 ------TTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCC---ACC 1542
Db      7002 ArgCysAlaSerArgSerArgArgArgCysAspSerProProArg------ProCysThr 7019

Qy      1543 TGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG---------AGCCCGTGG 1593
Db      7020 PheAlaAlaArgTrpTrpSerArgGlySerThrArgArgTrpTrpSerAlaAlaGlyTrp 7039

Qy      1594 CACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGT---------------- 1636
Db      7040 ArgSerAlaSerAlaGlyThrCysArgProThr-SerSerGlySerCysProProSerAl 7059

Qy      1637 ------------------------------GACGTGGCCAGATGGCAAGATGGTGAGCC 1665
Db      7059 aThrSerSerIleThrProAsnAlaSerArgAsnThrArgArgArgThrAlaMetArgPr 7079

Qy      1666 GGAA--------------------------CGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
Db      7079 oGluArgThrAspSerValThrHisValSerArgGlyArgArgGlyGluArgAlaAlaPr 7099

Qy      1699 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGC-------------- 1744
Db      7099 oAlaGlyPro------------------GluThrGlyLysProGluLysHisAlaSerTh 7113

Qy      1745 ----------------------------CCCACTGGAGTG-------------------- 1756
Db      7113 rIleValValGlyArgArgHisSerMetProSerGlyValProLeuLysArgProAspLe 7133

Qy      1757 ---------------------------------------------TGG------------ 1759
Db      7133 uThrAlaLeuThrAlaArgCysGlyIleArgArgSerTyrTrpTrpAlaGlyLeuAlaAr 7153

Qy      1760 ----------------------CCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGG 1797
Db      7153 gGlyCysGlyArgLeuArgIleProArgProAsnAsnTrpSerArgTrpProIleGluLe 7173

Qy      1798 ACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCC----------------------- 1834
Db      7173 uPheSerProMetTyrTrpIleProSerArgArgProGluTyrArgArgSerGlyLeuSe 7193

Qy      1835 ----TCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACA 1890
Db      7193 rSerGlyArgProArgArgArgSerArgLeuProTrpGluThrAlaArgSerSerAlaTr 7213

Qy      1891 AGAAGTGCAGTCGGGGCTACGAGCCCAACGA---GGATGGCACAGCCTGCGTGGGCTGGT 1947
Db      7213 pThrSerProThrSerSerArgThrProArgAlaGlyTrpProThr-------------- 7228

Qy      1948 GGAGCCCTGT---------GTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTC 1998
Db      7229 ----ProCysSerCysProValThrSerSerAlaThrThrThrSerSerCysThrTrpAl 7247

Qy      1999 CCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACC 2042
Db      7247 aThrThrThrSerTrpThrAlaSerSer-ThrSerTrpArgThr 7261
```

RESULT 15
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:

Pred. No.:	2.32e-10	Length:	19723
Score:	291.00	Matches:	220
Percent Similarity:	30.56%	Conservative:	48
Best Local Similarity:	25.09%	Mismatches:	297
Query Match:	7.04%	Indels:	313
DB:	15	Gaps:	48

US-09-914-958B-35 (1-2177) x US-10-084-846A-5 (1-19723)

```
Qy       1953 GGCTCCACCAGCCCACGC---AGGCTGTGCCATCCTCGTTGGGCTCGTAGCCCCGACTGC 1897
Db       6645 GlyAlaThrSerProArgProGlyProSerAlaProProTrpAla---ProProThrAla 6663

Qy       1896 ACTTCTTGTTGGTCCGGC---------------------------------------------- 1879
Db       6664 ThrAlaAlaTrpThrGlyArgArgSerSerArgThrGlyThrSerLysSerCysArgPro 6683

Qy       1879 ---------------------------------------------------------------- 1879
Db       6684 ThrProGlySerThrAlaArgCysGlyValProSerProGlyArgGluArgProSerAla 6703

Qy       1878 ------ACCTGTAGCTTCCATAGGTGTTGACACATACGGGCTTGTCTCGAGGGCACACGA 1825
Db       6704 AlaThrThrSerThrProThrArgCysTrpAlaArgProAlaAlaSerArgAlaAlaThr 6723

Qy       1824 ATGGGAACTGGATGCATTCATTGGTGT------------------------------------ 1798
Db       6724 CysGlyCysGlyThrAlaArgTrpThrThrThrAlaAlaArgThrArgAlaAlaThrPro 6743

Qy       1797 --------------------------CCATGC---------AATGGCCATTTTCCTGCT 1774
Db       6744 SerThrAlaTrpArgSerSerArgProProCysTrpArgSerArgSerTrpArgProAla 6763

Qy       1773 GGGAGAATCCTTGGCCACACTCCAGTGGGGCTGGGTCCTGAAGTGTGTCCTCATCCCGGG 1714
Db       6764 ProArgAlaLysSerAlaAlaProSerArgThrArgProThrArgCysArgArgValSer 6783

Qy       1713 GGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGG---------------------- 1675
Db       6784 GlyArgSerArgSer------SerAlaArgAlaArgTrpCysGlyTrpProProGlyArg 6801

Qy       1674 ------------------------------------------CCACGTTCCGGCTCACCA 1657
Db       6802 SerArgThrProProThrArgThrProProGlySerSerMetProArgProGlyArgPro 6821

Qy       1656 TCTTGCCATCTGGCCACGTCACCT------------CCACACTGCTGGCTTCATCCTTCC 1609
Db       6822 Arg---------ArgThrSerProValTrpArgArgGlyArgCysTrpThrArgAlaSer 6838

Qy       1608 CCAGGCCAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGC------------CTGAGC 1561
Db       6839 ThrGlySerAsnArgProThrProAlaTrpSerProGlyGlyTrpProArgTrpAlaTyr 6858

Qy       1560 CCCCGTCGATGATCCTCAGGTGGG---------------CCCCACTCTTCTTGGTGTAGA 1516
Db       6859 ProArgProValSerIleGlyAlaSerSerAspCysProProValSerValTyrCysThr 6878

Qy       1515 GCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTCGCAGCC 1456
Db       6879 LeuArgCysSerArgProThrProArgProGlyProArgCysTrpThrPro--------- 6895

Qy       1455 AGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTC 1396
Db       6896 ------CysArgSerArgArgProProSerTrpAlaSerArgSerSerAlaTrpSerCys 6913

Qy       1395 CATGGGACAAGATGAGGTCCAGCATCC------------CGTCTCCGTCGAAGTCGG--- 1351
Db       6914 ThrSerThrArg---SerProSerThrAlaArgHisArgArgArgArgArgSerArgSer 6932

Qy       1350 ---------------TCACCACACCCCCTGTGCCCCG---------------------- 1329
Db       6933 ArgSerIleProAlaSerPro-ThrProCysThrProTrpSerArgThrGlyArgAspAl 6952

Qy       1328 ----------GCCCTCAGGCTCCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGT 1279
Db       6952 aArgTrpArgAla-CysAlaAlaCysAlaThrArgTrpSerSerSerArgProArgArgT 6972

Qy       1278 CTCCGTGCTCTCTACGGATGACGCGGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAA 1219
Db       6972 hrProThrSerArgAlaAlaThrArgThrArgThr-TrpCysArgSer---------Arg 6988

Qy       1218 TGTTGTTGAAGAAGATCTCCAGCTCCTGGT---------CATTGTCAAAGTCGGCGGTGA 1168
Db       6989 ValCysArgArgArgSerCysValProGlyCysSerThrArgCysAlaSerArgSer--- 7007

Qy       1167 TGACCGTGCGGACAGGGGAGGGCATGGAGAACTTGGGTGAGGCGATGTCCCGGAAGCGGA 1108
Db       7008 ------------------------------------ArgArgArgCys------AspSer 7013

Qy       1107 CCTTCCCATGGGTGCTCATTTGCAGATAG----------------------AGGCGGTGG 1070
Db       7014 ProProArgProCysThrPheAlaAlaArgTrpTrpSerArgGlySerThrArgArgTrp 7033

Qy       1069 GGGCCATTCCAGTTGCCATAGACGATGTCCACTTTGCCATCACGGTTGAAGTCAGCCAGG 1010
Db       7034 ------------------------TrpSerAlaAlaGlyTrpArgSerAlaSerAla--- 7044

Qy       1009 GCGACACCTCGCCCATGCTGGTGGGGGTCGTCCACACCAGCACTGGCCGCAGCGTCCACA 950
Db       7045 GlyThrCysArgProThrSerSerGlySerCysProProSerAlaThrSerSerIleThr 7064

Qy        949 AAGGTGCCATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAG 890
Db       7065 ProAsnAlaSerArgAsnThrArgArgArgThrAlaMetArgProGluArgThrAspSer 7084

Qy        889 ATATCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCT------ 836
Db       7085 ValThrHisValSerArgGlyArgArgGlyGluArgAlaAlaProAlaGlyProGluThr 7104

Qy        836 ---------------------------------------------------------- 836
Db       7105 GlyLysProGluLysHisAlaSerThrIleValValGlyArgArgHisSerMetProSer 7124

Qy        835 ---GTATATTTGCTGACCCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGAATGC----- 784
Db       7125 GlyValProLeuLysArgPro------AspLeuThrAlaLeuThrAlaArg-CysGlyIl 7142

Qy        783 ------------------------------------------------------CCCGG 779
Db       7142 eArgArgSerTyrTrpTrpAlaGlyLeuAlaArgGlyCysGlyArgLeuArgIleProAr 7162

Qy        778 GAGAGGTCACTGGCCTCAGGGTCCATTTCAA--------------------TGAGGGCA 740
Db       7162 gProAsnAsnTrpSerArgTrpProIleGluLeuPheSerProMetTyrTrpIleProSe 7182

Qy        739 TCAGGGCCCACATTACCGTAGGCGTAAT------------------------------ 712
Db       7182 rArgArgProGluTyrArgArgSerGlyLeuSerSerGlyArgProArgArgArgSerAr 7202

Qy        711 -------TGGCAA---TGTAGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCACACAGGCC 662
Db       7202 gLeuProTrpGluThrAlaArgSerSerAlaTrpThrSerProThrSerSerArgThrPr 7222

Qy        661 ACAGAGCGTCCGGCAAAGAGGCTGGCCACACCACGGGCCACGTTGACCTCATCGCTCAGG 602
Db       7222 oArgAla------------GlyTrpPro--------------------------ThrPr 7229

Qy        601 ATGT---CTTCCCACCGGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGG----- 550
Db       7229 oCysSerCysProValThrSerSerAlaThr---ThrThrSerSerCysThrTrpAlaTh 7248

Qy        549 ------------------------------------------------------CCACCCCC 542
Db       7248 rThrThrSerTrpThrAlaSerSerThrSerTrpArgThrSerAlaAlaThrSerProPr 7268

Qy        541 GAGAAGGCATTATTGGTGTTGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGTCGATGTCG 482
```

```
                ||||||||                                  ||||||   |||   |||
Db       7268 oArgArg---------------------SerCysSerProGlySerProThrArgAlaGl 7281

Qy        481 CAGGCTGTGACCC--------------------------CAATGGCGTTCCCCTGCCGG 449
                    |||                               ||||||||||||
Db       7281 ySerAlaTrpProAsnLeuThrArgThrAlaGlyProArgSerTrpArgSerProArgTh 7301

Qy        448 TCCCGCAGCGCGTAGTAGGGTGAGCTGC---------GCTCATCGACCG------CGATG 404
                |||   |||      |||   |||           :::   ||||||      |||
Db       7301 rProGlyAlaThrTrpArgSerSerAlaCysThrProSerAlaArgProCysThrArgPr 7321

Qy        403 TTCACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCATTGTACCCC 344
                ||||||:::      ||||||   ||||||          |||         |||||
Db       7321 oTrpProAlaSerGlyArgProGlyGlyThrSer---------TrpArgSerProThrPr 7338

Qy        343 GCCACGACGATCTCAAAGTCCCCATCATGGTCCACATCAGTAACTGCCACACCATAGTTG 284
                |         |||      |||      |||                |||      :::
Db       7338 o------CysSerGlyTrpProThrThrGlyThrArgSer---AlaProArgSerThrGl 7355

Qy        283 AGCTGGGTGGGATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTGAAC 224
                |||               |||:::   :::      ||||||
Db       7355 yThrGlyArgThr---------SerArgAlaSerProThrSerTrpArgThrGlyThrCy 7372

Qy        223 ATGGGTT------CAGCCCGCTGGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGC 170
                |||         :::||||||   |||   :::   |||      ||||||   |||
Db       7372 sTrpSerProTrpArgProAlaSerThrAlaArg---Trp------ThrArgProAlaAs 7389

Qy        169 AGGAACGGTAACATCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCG 110
                      ||||||      ||||||:::||||||         :::   |||
Db       7389 nTrpTrpAlaGluSerTrpTrpArgArgGlyArgArgSerProAlaProAlaSerServa 7409

Qy        109 GCCCCGCCGCCTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCGCCTCTGCCGCCGGCGCCG 50
                |||                   :::   ||||||   |||   ||||||
Db       7409 lPro-----------------ProSerSerAlaArgThrArgTrpCysArgThrProTh 7423

Qy         49 CCGCCTGCTTGCTCCCAGCCCCGGTCCCGGGCTGGCCTCGAGC 7
                :::      |||||||||   |||      |||::::::
Db       7423 rSerVal------ProSerPro---ProSerArgAlaThrAla 7434


Search completed: August  3, 2004, 20:08:23
Job time : 221 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 12:54:35 ; Search time 8286 Seconds
(without alignments)
11387.616 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2177	100.0	2177	6	BD269644	BD269644 Human sec
2	2043.8	93.9	2089	6	AX357116	AX357116 Sequence
3	1940.4	89.1	2681	9	HSA276171	AJ276171 Homo sapi
4	1934	88.8	2889	9	AK057190	AK057190 Homo sapi
5	1913.4	87.9	2413	6	AX427468	AX427468 Sequence
6	1893	87.0	2263	6	AX202100	AX202100 Sequence
7	1808.8	83.1	2589	6	AX357100	AX357100 Sequence
8	1807.8	83.0	2039	6	AX357115	AX357115 Sequence
9	1807.8	83.0	2145	6	AX357117	AX357117 Sequence
10	1794.4	82.4	1959	6	AX470036	AX470036 Sequence
11	1794.4	82.4	2507	6	AX470035	AX470035 Sequence
12	1794.4	82.4	2507	9	HSA279016	AJ279016 Homo sapi
13	1793.4	82.4	1947	6	AX470056	AX470056 Sequence
14	1793.4	82.4	1947	9	HSA421515	AJ421515 Homo sapi
15	1476.4	67.8	1973	10	MMU421516	AJ421516 Mus muscu
16	1438	66.1	2147	9	BC034245	BC034245 Homo sapi
17	1408.8	64.7	2178	6	AX876177	AX876177 Sequence
18	1408.8	64.7	2178	6	BD155992	BD155992 Primer fo
19	1408.8	64.7	2178	9	AK001182	AK001182 Homo sapi
20	1345.8	61.8	1501	6	AR339541	AR339541 Sequence
21	1137.8	52.3	1558	10	BC024472	BC024472 Mus muscu
22	1025.6	47.1	1143	6	BD205640	BD205640 97 human
23	788	36.2	789	6	AR379939	AR379939 Sequence
24	556.8	25.6	608	6	AX867169	AX867169 Sequence
25	556.8	25.6	608	6	BD147231	BD147231 Primer fo
26	449	20.6	700	10	RNU78304	U78304 Rattus norv
c 27	229	10.5	492	11	G27517	G27517 human STS S
c 28	225	10.3	146124	9	AL358938	AL358938 Human DNA
c 29	207	9.5	161235	9	AL139239	AL139239 Human DNA
c 30	207	9.5	172383	2	AC015630	AC015630 Homo sapi
31	174	8.0	300750	1	AP006576	AP006576 Gloeobact
c 32	164	7.5	163348	10	AC119236	AC119236 Mus muscu
33	164	7.5	243770	2	AC120763	AC120763 Rattus no
c 34	164	7.5	252404	2	AC097117	AC097117 Rattus no
c 35	161.6	7.4	395	11	G25373	G25373 human STS E
36	161.2	7.4	172383	2	AC015630	AC015630 Homo sapi
c 37	158	7.3	164766	10	AC144792	AC144792 Mus muscu
38	148.2	6.8	409	11	HSJ123C5	AL159715 STS from
c 39	145.4	6.7	210385	10	AC106128	AC106128 Rattus no
40	142.2	6.5	189196	10	AL603804	AL603804 Mouse DNA
41	130.2	6.0	229480	5	AL929150	AL929150 Zebrafish
c 42	114.8	5.3	110000	2	AC099306_5	Continuation (6 of
43	110.4	5.1	743	11	BV022941	BV022941 S212P6878
44	109	5.0	707	6	AR379664	AR379664 Sequence
c 45	101.4	4.7	166810	2	BX005216	BX005216 Danio rer

ALIGNMENTS

RESULT 1
BD269644
LOCUS BD269644 2177 bp DNA linear PAT 17-JUL-2003
DEFINITION Human secretory proteins.
ACCESSION BD269644
VERSION BD269644.1 GI:33079412
KEYWORDS JP 2002537805-A/13.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2177)
AUTHORS Tang,T.Y., Lal,P., Baughn,M.R., Yue,H., Young,J.A., Lu,D.A.M. and Azimzai,Y.
TITLE Human secretory proteins

```
  JOURNAL   Patent: JP 2002537805-A 13 12-NOV-2002;
            INCYTE PHARMACEUTICALS INC
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002537805-A/13
            PD   12-NOV-2002
            PF   03-MAR-2000 JP 2000602763
            PR   05-MAR-1999 US    60/123117
            PI   TOM Y TANG,PREETI LAL,MARIAH R BAUGHN,HENRY YUE,JANICE AU  PI
             YOUNG,
            PI   DYUNG AINA M LU,YALDA AZIMZAI
            PC   C12N15/09,A01K67/027,A61K38/00,A61K38/22,A61K45/00,A61P1/00,
            PC   A61P9/00,
            PC   A61P25/00,A61P29/00,A61P31/00,A61P35/00,C07K14/47,C07K16/18,
            PC   C12N1/15,
            PC   C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00 PC
            ,C12N5/00,A61K37/02,
            PC   A61K37/24
            CC   Incyte ID No: 2991370CB1
            FH   Key             Location/Qualifiers
            FT   source          1. .2177
            FT                   /organism='Homo sapiens (human)'.
FEATURES             Location/Qualifiers
     source          1. .2177
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 2177;  DB 6;  Length 2177;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2177;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560

Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Qy       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980

Qy       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040

Qy       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100

Qy       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160

Qy       2161 TCTTGAAAAAAAAAAAA 2177
             |||||||||||||||||
Db       2161 TCTTGAAAAAAAAAAAA 2177
```

RESULT 2
AX357116
LOCUS AX357116 2089 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 17 from Patent WO0206478.
ACCESSION AX357116
VERSION AX357116.1 GI:18674296
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
TITLE Extracellular matrix protein
JOURNAL Patent: WO 0206478-A 17 24-JAN-2002;
Akzo Nobel N.V. (NL)
FEATURES Location/Qualifiers
source 1. .2089
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 93.9%; Score 2043.8; DB 6; Length 2089;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2045; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy        191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy        311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy        371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy        431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy        491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy        551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy        611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy        671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy        731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy        791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy        851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy        911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy        971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy       1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy       1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy       1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302
```

```
Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy       1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy       1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy       1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy       1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy       1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy       1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy       1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy       1931 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1902

Qy       1991 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 2050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 1962

Qy       2051 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2022

Qy       2111 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2082

Qy       2171 AAAAAAA 2177
             |||||||
Db       2083 AAAAAAA 2089
```

```
RESULT 3
HSA276171
LOCUS       HSA276171              2681 bp    mRNA    linear   PRI 20-JUL-2000
DEFINITION  Homo sapiens mRNA for ASPIC (acidic secreted protein in cartilage)
            (ASPIC1 gene).
ACCESSION   AJ276171
VERSION     AJ276171.1  GI:9368805
KEYWORDS    acidic secreted protein in cartilage; ASPIC1 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bolton,M.C., Wait,R. and Saklatvala,J.
  TITLE     Cloning of ASPIC, a novel protein secreted by human normal and
            osteoarthritic cartilage, identified by 2D electrophoresis and mass
            spectrometry
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2681)
  AUTHORS   Bolton,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2000) Bolton M.C., Cell signalling, Kennedy
            Institute of Rheumatology, 1 Aspenlea Road, Hammersmith, London W6
            8LH, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2681
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     gene            1. .2681
                     /gene="ASPIC1"
     5'UTR           1. .149
                     /gene="ASPIC1"
                     /note="A second transcription start site at +31"
                     /evidence=experimental
     CDS             150. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98267.1"
                     /db_xref="GI:9368806"
                     /db_xref="GOA:Q9NQ80"
                     /db_xref="SPTREMBL:Q9NQ80"
                     /translation="MAPSADPGMSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLP
                     PDYDSNPTQLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSS
                     PYYALRDRQGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDI
                     LSDEVNVARGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDL
                     SRGILALRDVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTF
                     VDAAASAGVDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDI
                     ASPKFSMPSPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEEL
                     NPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVP
                     RTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWP
                     DGKMVSRNVASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQF
                     PFVCPRDKPVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAA
                     ATAAAAAAAGAATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     150. .230
                     /gene="ASPIC1"
     CDS             174. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98268.1"
                     /db_xref="GI:9368807"
                     /db_xref="GOA:Q9NQ79"
                     /db_xref="SPTREMBL:Q9NQ79"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAA
                     AGAATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     174. .230
                     /gene="ASPIC1"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
```

```
     CDS             183. .2135
                     /gene="ASPIC1"
                     /standard_name="acidic secreted protein in cartilage"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
                     /codon_start=1
                     /evidence=experimental
                     /product="ASPIC"
                     /protein_id="CAB98269.1"
                     /db_xref="GI:9368808"
                     /db_xref="GOA:Q9NQ78"
                     /db_xref="SPTREMBL:Q9NQ78"
                     /translation="MLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPTQLN
                     YGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDRQGN
                     AIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVARGV
                     ASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALRDVA
                     AEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAGVDD
                     PHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMPSPV
                     RTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEPEGR
                     GTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFARGA
                     KVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRNVAS
                     GEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDKPVC
                     VNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAAAGA
                     ATAAPVLVDGDLNLGSVVKESCEPSC"
     sig_peptide     183. .230
                     /gene="ASPIC1"
                     /note="putative alternative start, less identity to the
                     Kozak rule"
     mat_peptide     231. .2132
                     /gene="ASPIC1"
                     /product="ASPIC"
                     /standard_name="acidic secreted protein in cartilage"
     3'UTR           2136. .2681
                     /gene="ASPIC1"
     polyA_signal    2642. .2647
                     /gene="ASPIC1"
ORIGIN

  Query Match           89.1%;  Score 1940.4;  DB 9;  Length 2681;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1941;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         28 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 87

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 147

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 207

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 267

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 327

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 387

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 447

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        448 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCT 507

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        508 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 567

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        568 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 627

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        628 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 687

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        688 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 747

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        748 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 807

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        808 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 867

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        868 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 927

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 987

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1047

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1048 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1107

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1167

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1227

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1228 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1287

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1347

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1407

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1467

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1527

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1587
```

```
Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1647

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1648 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1707

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1708 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1767

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1768 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1827

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1828 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1887

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1888 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1947

Qy       1921 AGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||
Db       1948 AGGATGGCACAGCCTGCGTGGG 1969
```

RESULT 4
AK057190

```
LOCUS       AK057190               2889 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ32628 fis, clone SYNOV1000118, highly similar
            to Homo sapiens mRNA for ASPIC (acidic secreted protein in
            cartilage) (ASPIC1 gene).
ACCESSION   AK057190
VERSION     AK057190.1  GI:16552786
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
            Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
            Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
            Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
            Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
            Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
            Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2889)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1. .2889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SYNOV1000118"
                     /tissue_type="synovial membrane tissue from rheumatoid
                     arthritis"
                     /clone_lib="SYNOV1"
                     /note="cloning vector: pME18SFL3"
ORIGIN
```

```
  Query Match             88.8%;  Score 1934;  DB 9;  Length 2889;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1937;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 294

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 354

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        355 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 414

Qy        181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 474

Qy        241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 534

Qy        301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 594

Qy        361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 654

Qy        421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCT 714

Qy        481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 774

Qy        541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 834

Qy        601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 894

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 954

Qy        721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 1014

Qy        781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 1074

Qy        841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 GCCGAGGCGTCAGTGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 1134
```

```
Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 1194

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1195 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1254

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1314

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCTAGTTCT 1374

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1434

Qy       1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1494

Qy       1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1554

Qy       1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1614

Qy       1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1615 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1674

Qy       1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1675 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1734

Qy       1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1735 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1794

Qy       1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1795 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1854

Qy       1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
              |||||||||||||||||||| ||||||||||||||||||||||||| |||||||||||||||
Db       1855 CCAGCAGTGTGGAGGTGACGTGGTCAGATGGCAAGATGGTGAGCCGGAACGCGGCCAGCG 1914

Qy       1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1915 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1974

Qy       1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1975 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 2034

Qy       1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2035 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 2094

Qy       1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2095 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 2154

Qy       1921 AGGATGGCACAGCCTGCGTGGG 1942
              |||||||||||||||||||||||
Db       2155 AGGATGGCACAGCCTGCGTGGG 2176
```

```
RESULT 5
AX427468
LOCUS       AX427468                2413 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 41 from Patent WO0123558.
ACCESSION   AX427468
VERSION     AX427468.1  GI:21537614
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hodgson,D.M., Lincoln,S.E., Russo,F.D., Spiro,P.A., Banville,S.C.,
            Bratcher, Shawn,R., Dufour,G.F., Cohen,H.J., Rosen,B.H., Shah,P.,
            Chalup,M.S., Hillman,J.L., Jones,A.L., Yu,J.Y., Greenawalt,L.B.,
            Panzer,S.R., Roseberry,A.M., Wright,R.J., Chen,W., Liu,T.,
            Yap,P.E., Stockdreher,T.K., Amshey,S. and Fong,W.T.
  TITLE     Human secretory molecules
  JOURNAL   Patent: WO 0123558-A 41 05-APR-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2413
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 481454.4.dec"
ORIGIN
```

Query Match 87.9%; Score 1913.4; DB 6; Length 2413;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1917; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy          19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         318 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 377

Qy          79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         378 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 437

Qy         139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         438 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 497

Qy         199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         498 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 557

Qy         259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         558 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 617

Qy         319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         618 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 677

Qy         379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         678 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 737

Qy         439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         738 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 797

Qy         499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         798 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 857

Qy         559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         858 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 917
```

```
Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        918 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGATGCTCTGTGGCCTGTGTGGACAGAA 977

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 1037

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 1097

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 1157

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1217

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1277

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1337

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1397

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1457

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1517

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1577

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1637

Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1697

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1757

Qy       1459 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1817

Qy       1519 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1877

Qy       1579 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1638
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1878 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1937

Qy       1639 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1997

Qy       1699 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 1758
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 2057

Qy       1759 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 1818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 2117

Qy       1819 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 1878
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2118 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 2177

Qy       1879 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 1938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2178 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 2237

Qy       1939 TGG 1941
              |||
Db       2238 TGG 2240
```

```
RESULT 6
AX202100
LOCUS       AX202100                2263 bp    DNA     linear   PAT 30-AUG-2001
DEFINITION  Sequence 30 from Patent WO0153531.
ACCESSION   AX202100
VERSION     AX202100.1  GI:15391887
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Phippard,D., Vasanthakamur,G., Dotson,S. and Ma,X.J.
  TITLE     Osteoarthritis tissue derived nucleic acids, polypeptides, vectors,
            and cells
  JOURNAL   Patent: WO 0153531-A 30 26-JUL-2001;
            Pharmacia Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .2263
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             87.0%;  Score 1893;  DB 6;  Length 2263;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy          19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 212

Qy          79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 272

Qy         139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 332

Qy         199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 392

Qy         259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 452

Qy         319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 512

Qy         379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
```

```
Db        513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 572
Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
Db        573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 632
Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
Db        633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 692
Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
Db        693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 752
Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
Db        753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 812
Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
Db        813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 872
Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 798
Db        873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 932
Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
Db        933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 992
Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
Db        993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1052
Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
Db       1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1112
Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
Db       1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1172
Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
Db       1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1232
Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
Db       1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1292
Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
Db       1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1352
Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
Db       1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1412
Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
Db       1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1472
Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
Db       1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1532
Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
Db       1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1592
Qy       1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1517
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652
Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577
Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712
Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637
Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772
Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697
Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832
Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1757
Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1892
Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816
Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952
Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012
Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936
Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072
Qy       1937 CGTGG 1941
Db       2073 CGTGG 2077
```

```
RESULT 7
AX357100
LOCUS       AX357100               2589 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 1 from Patent WO0206478.
ACCESSION   AX357100
VERSION     AX357100.1  GI:18674281
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 1 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2589
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match           83.1%;  Score 1808.8;  DB 6;  Length 2589;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1810;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
Db         43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102
Qy        191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162
Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222
```

```
Qy      311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
Db      223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy      371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
Db      283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy      431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
Db      343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy      491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
Db      403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy      551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
Db      463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy      611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db      523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy      671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db      583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy      731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
Db      643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy      791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db      703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy      851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db      763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy      911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db      823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy      971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db      883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy     1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db      943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy     1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db     1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy     1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db     1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy     1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db     1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy     1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db     1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy     1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db     1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy     1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db     1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy     1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db     1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy     1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db     1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy     1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db     1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy     1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db     1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy     1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db     1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy     1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db     1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy     1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db     1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy     1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db     1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy     1931 AGCCTGCGTGGG 1942
Db     1843 AGCCTGCGTGGG 1854
```

```
RESULT 8
AX357115
LOCUS       AX357115               2039 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 16 from Patent WO0206478.
ACCESSION   AX357115
VERSION     AX357115.1  GI:18674295
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 16 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2039
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             83.0%;  Score 1807.8;  DB 6;  Length 2039;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
Db       43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy      191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
```

```
Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162
Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222
Qy        311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
Db        223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282
Qy        371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
Db        283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342
Qy        431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
Db        343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402
Qy        491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
Db        403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462
Qy        551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
Db        463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522
Qy        611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db        523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582
Qy        671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db        583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642
Qy        731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCT 790
Db        643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCT 702
Qy        791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db        703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762
Qy        851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db        763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822
Qy        911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db        823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882
Qy        971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db        883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942
Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db        943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002
Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062
Qy       1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122
Qy       1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182
Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
```

```
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242
Qy       1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302
Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362
Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422
Qy       1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482
Qy       1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542
Qy       1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602
Qy       1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662
Qy       1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722
Qy       1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782
Qy       1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842
Qy       1931 AGCCTGCGTGG 1941
Db       1843 AGCCTGCGTGG 1853
```

```
RESULT 9
AX357117
LOCUS       AX357117               2145 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 18 from Patent WO0206478.
ACCESSION   AX357117
VERSION     AX357117.1  GI:18674297
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Heus,H.C., Nelissen,R.L. and Meeuwisse,C.M.
  TITLE     Extracellular matrix protein
  JOURNAL   Patent: WO 0206478-A 18 24-JAN-2002;
            Akzo Nobel N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .2145
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             83.0%;  Score 1807.8;  DB 6;  Length 2145;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy      131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
Db       43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy      191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
Db      103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy      251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
Db      163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy      311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
Db      223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy      371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
Db      283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy      431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
Db      343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy      491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
Db      403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy      551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
Db      463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy      611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db      523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy      671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db      583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy      731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
Db      643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy      791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db      703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy      851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db      763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy      911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db      823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy      971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db      883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy     1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db      943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy     1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db     1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy     1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db     1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy     1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db     1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy     1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db     1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy     1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db     1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy     1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db     1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy     1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db     1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy     1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db     1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy     1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db     1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy     1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db     1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy     1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db     1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy     1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db     1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy     1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db     1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy     1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db     1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy     1931 AGCCTGCGTGG 1941
Db     1843 AGCCTGCGTGG 1853
```

```
RESULT 10
AX470036
LOCUS       AX470036               1959 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 2 from Patent WO02053709.
ACCESSION   AX470036
VERSION     AX470036.1  GI:22205283
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Richter,W. and Steck,E.
  TITLE     Cartilage marker gene, method for the production and use thereof
  JOURNAL   Patent: WO 02053709-A 2 11-JUL-2002;
            CYTONET GMBH & CO KG (DE)
FEATURES             Location/Qualifiers
     source          1. .1959
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
```

/db_xref="taxon:9606"

ORIGIN

Query Match 82.4%; Score 1794.4; DB 6; Length 1959;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 11
AX470035
LOCUS AX470035 2507 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 1 from Patent WO02053709.
ACCESSION AX470035
VERSION AX470035.1 GI:22205282
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Richter,W. and Steck,E.
TITLE Cartilage marker gene, method for the production and use thereof
JOURNAL Patent: WO 02053709-A 1 11-JUL-2002;
CYTONET GMBH & CO KG (DE)
FEATURES Location/Qualifiers
source 1. .2507

```
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             82.4%;  Score 1794.4;  DB 6;  Length 2507;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796


RESULT 12
HSA279016
LOCUS       HSA279016               2507 bp    mRNA    linear   PRI 11-MAR-2001
DEFINITION  Homo sapiens mRNA for chondrocyte expressed protein 68 kDa (CEP-68
            gene).
ACCESSION   AJ279016
VERSION     AJ279016.1  GI:10178882
KEYWORDS    CEP-68 gene; chondrocyte expressed protein 68 kDa.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Steck,E., Benz,K., Lorenz,H., Loew,M., Gress,T. and Richter,W.
  TITLE     Chondrocyte expressed protein-68 (CEP-68), a novel human marker
            gene for cultured chondrocytes
```

```
   JOURNAL   Biochem. J. 353 (Pt 2), 169-174 (2001)
   MEDLINE   21066088
    PUBMED   11139377
REFERENCE    2  (bases 1 to 2507)
  AUTHORS    Steck,E.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-SEP-2000) Steck E., Research, Department of
             Orthopaedic Surgery, University of Heidelberg, Schlierbacher
             Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES             Location/Qualifiers
     source          1. .2507
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="chondrocyte"
                     /tissue_type="cartilage"
     gene            1. .2507
                     /gene="CEP-68"
     CDS             1. .1962
                     /gene="CEP-68"
                     /function="unknown"
                     /note="protein with EGF-like calcium binding motif"
                     /codon_start=1
                     /product="chondrocyte expressed protein 68 kDa"
                     /protein_id="CAC08451.1"
                     /db_xref="GI:10178883"
                     /db_xref="GOA:Q9NQ79"
                     /db_xref="SPTREMBL:Q9NQ79"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVGTLGQSPGPRPTTPTAAAATAAAAAA
                     AGAATAAPVLVDGDLNLGSVVKESCEPSC"
     polyA_signal    2469. .2474
                     /gene="CEP-68"
                     /evidence=experimental
ORIGIN

  Query Match           82.4%;  Score 1794.4;  DB 9;  Length 2507;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
```

```
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706

Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766

Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826

Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886

Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942

Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

```
RESULT 13
AX470056
LOCUS       AX470056                1947 bp    DNA     linear   PAT 09-AUG-2002
DEFINITION  Sequence 22 from Patent WO02053709.
ACCESSION   AX470056
VERSION     AX470056.1  GI:22205301
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Richter,W. and Steck,E.
  TITLE     Cartilage marker gene, method for the production and use thereof
  JOURNAL   Patent: WO 02053709-A 22 11-JUL-2002;
            CYTONET GMBH & CO KG (DE)
FEATURES             Location/Qualifiers
     source          1. .1947
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             82.4%;  Score 1793.4;  DB 6;  Length 1947;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1794;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206

Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266

Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326

Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386

Db       181 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446

Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506

Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566

Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626

Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686

Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746

Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806

Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866

Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926

Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986

Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046

Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106

Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166

Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226

Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286

Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346

Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406

Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466

Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526

Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586

Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
```

```
Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

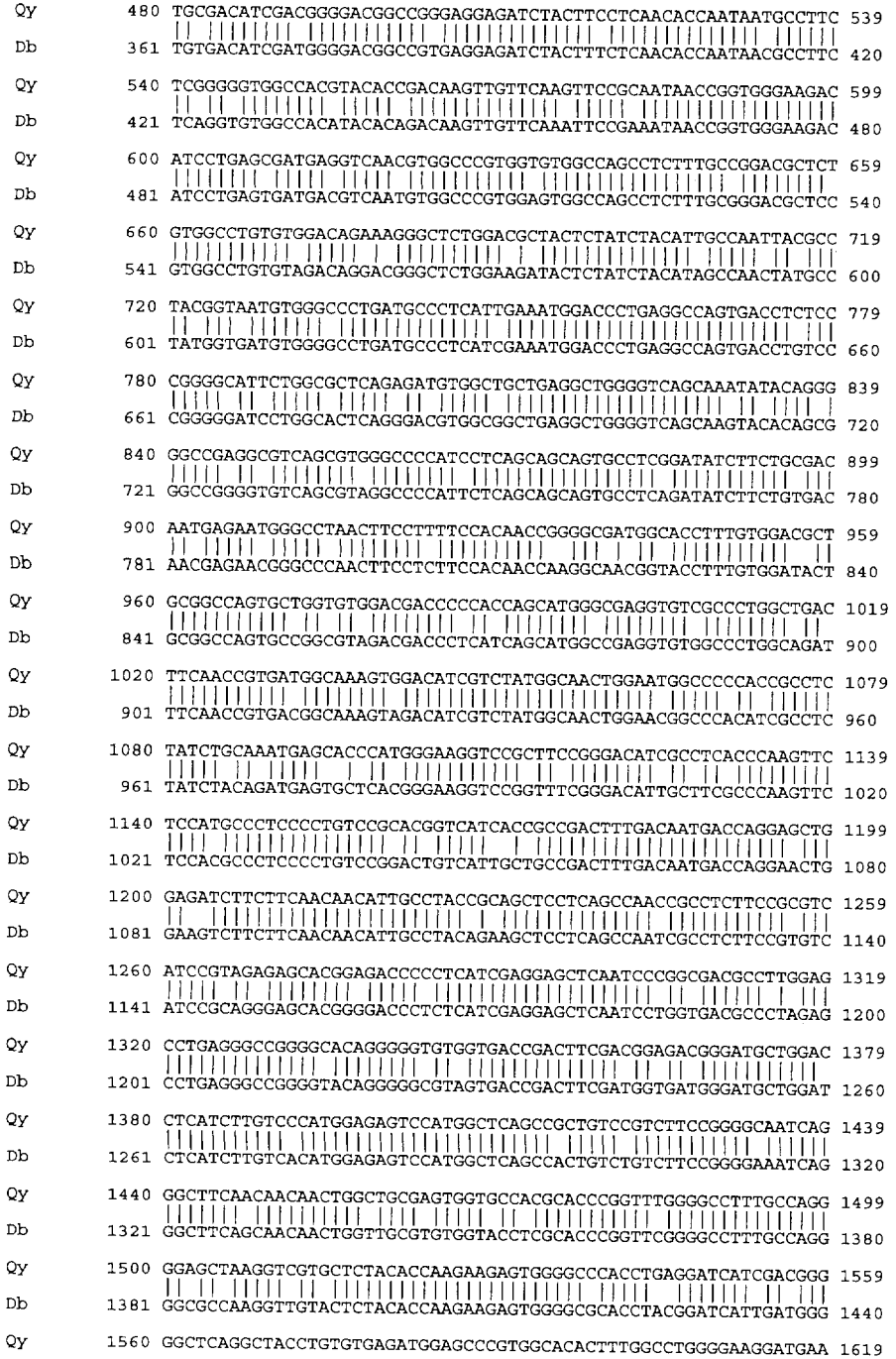
```
RESULT 14
HSA421515
LOCUS       HSA421515               1947 bp    mRNA    linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens mRNA for CRTAC1-B protein (CRTAC1 gene).
ACCESSION   AJ421515
VERSION     AJ421515.1  GI:19171210
KEYWORDS    CRTAC1 gene; CRTAC1-B protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Steck,E. and Richter,W.
  TITLE     A novel form of CRTAC1 expressed in human and mouse brain generated
            by alternative splicing of a newly identified last exon
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1947)
  AUTHORS   Steck,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Steck E., Research, Department of
            Orthopaedic Surgery, University of Heidelberg, Schlierbacher
            Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1947
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .1947
                     /gene="CRTAC1"
     CDS             1. .1914
                     /gene="CRTAC1"
                     /codon_start=1
                     /evidence=experimental
                     /product="CRTAC1-B protein"
                     /protein_id="CAD13394.1"
                     /db_xref="GI:19171211"
                     /db_xref="GOA:Q8TE52"
                     /db_xref="SPTREMBL:Q8TE52"
                     /translation="MSRMLPFLLLLWFLPITEGSQRAEPMFTAVTNSVLPPDYDSNPT
                     QLNYGVAVTDVDHDGDFEIVVAGYNGPNLVLKYDRAQKRLVNIAVDERSSPYYALRDR
                     QGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWEDILSDEVNVA
                     RGVASLFAGRSVACVDRKGSGRYSIYIANYAYGNVGPDALIEMDPEASDLSRGILALR
                     DVAAEAGVSKYTGGRGVSVGPILSSSASDIFCDNENGPNFLFHNRGDGTFVDAAASAG
                     VDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSTHGKVRFRDIASPKFSMP
                     SPVRTVITADFDNDQELEIFFNNIAYRSSSANRLFRVIRREHGDPLIEELNPGDALEP
                     EGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFNNNWLRVVPRTRFGAFA
                     RGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGKDEASSVEVTWPDGKMVSRN
                     VASGEMNSVLEILYPRDEDTLQDPAPLECGQGFSQQENGHCMDTNECIQFPFVCPRDK
                     PVCVNTYGSYRCRTNKKCSRGYEPNEDGTACVAQVAFLGGYSSAASRISEPLSRASYL
                     SLGLGLCLQLYAL"
     polyA_signal    1923. .1928
                     /gene="CRTAC1"
ORIGIN
```

```
  Query Match             82.4%;  Score 1793.4;  DB 9;  Length 1947;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1794;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900
```

```
Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

```
RESULT 15
MMU421516
LOCUS       MMU421516               1973 bp    mRNA    linear   ROD 01-MAR-2002
DEFINITION  Mus musculus mRNA for CRTAC1-B protein (CRTAC1 gene).
ACCESSION   AJ421516
VERSION     AJ421516.1  GI:19171212
KEYWORDS    CRTAC1 gene; CRTAC1-B protein.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Steck,E. and Richter,W.
  TITLE     A novel form of CRTAC1 expressed in human and mouse brain generated
            by alternative splicing of a newly identified last exon
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1973)
  AUTHORS   Steck,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Steck E., Research, Department of
            Orthopaedic Surgery, University of Heidelberg, Schlierbacher
            Landstr. 200a, Heidelberg, 69118, GERMANY
FEATURES             Location/Qualifiers
     source          1. .1973
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .1973
                     /gene="CRTAC1"
     CDS             1. .1941
                     /gene="CRTAC1"
                     /codon_start=1
                     /evidence=experimental
                     /product="CRTAC1-B protein"
                     /protein_id="CAD13395.1"
                     /db_xref="GI:19171213"
                     /db_xref="GOA:Q8R555"
                     /db_xref="SPTREMBL:Q8R555"
                     /translation="MAPSADPGMVRMALLLLPPLWLLPLTGGSQRAEPMFTAVTNSVL
                     PPDYDSNPTQLNYGVAVTDVDHDGDFEIVVAGYTGPNLVLKYNRAQNRLVNIAVDERS
                     SPYYALRDRQGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWED
                     ILSDDVNVARGVASLFAGRSVACVDRTGSGRYSIYIANYAYGDVGPDALIEMDPEASD
                     LSRGILALRDVAAEAGVSKYTAGRGVSVGPILSSSASDIFCDNENGPNFLFHNQGNGT
                     FVDTAASAGVDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSAHGKVRFRD
                     IASPKFSTPSPVRTVIAADFDNDQELEVFFNNIAYRSSSANRLFRVIRREHGDPLIEE
                     LNPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFSNNWLRVV
                     PRTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGRDEASSVEVTW
                     PDGKMVSRSVASEEMNSVLEILYPQDEDKLQNTAPLECGQGFSQQDNGHCMDTNECIQ
                     FPFVCPRDKPVCVNTYGSYRCRTNKRCNRGYEPNEDGTACVAQVAFLGGYSSAAFRLS
                     EPLSQASYLSLGLGLCLQLYAL"
     polyA_signal    1950. .1955
                     /gene="CRTAC1"
ORIGIN
```

```
  Query Match             67.8%;  Score 1476.4;  DB 10;  Length 1973;
  Best Local Similarity   88.5%;  Pred. No. 1.2e-275;
  Matches 1613;  Conservative    0;  Mismatches  206;  Indels    3;  Gaps    1;

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTC--- 179
              |||||||||||||||||||| |||||| |||||||     | ||||||||| || |
Db          1 ATGGCTCCGAGCGCTGACCCGGGCATGGTCAGGATGGCTTTGCTCCTGCTGCCGCCCCTG 60

Qy        180 TGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACC 239
              ||| |||||||  |||||| ||| |||||| ||||||| ||||||||||||||||||||||
Db         61 TGGCTTCTGCCTCTCACTGGGGGATCCCAGAGGGCTGAGCCCATGTTCACTGCAGTCACC 120

Qy        240 AACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCA 299
              ||||||||  |||||  |||||||||||||||  || ||||||||||||||||||| ||||||
Db        121 AACTCAGTCCTGCCCCCTGACTATGACAGCAACCCCACCCAGCTCAACTATGGAGTGGCA 180

Qy        300 GTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCC 359
              || || |||||||||||||||||||||||| |||||||| |||||||||||||  || |||
Db        181 GTGACAGATGTGGACCATGATGGGGACTTCGAGATCGTTGTGGCGGGGTACACCGGCCCC 240

Qy        360 AACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAG 419
              ||||||||||||||||  | || |||||||| || |||||||| || ||||| || |||
Db        241 AACCTGGTTCTGAAGTACAATCGAGCCCAGAATCGACTGGTGAATATTGCGGTGGACGAG 300

Qy        420 CGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCC 479
              |||||||||||||||| || ||| ||||||| |||||||||||||| ||||||||||||
Db        301 CGCAGCTCACCCTACTATGCTCTGAGGGACCGTCAGGGGAACGCCATCGGGGTCACAGCC 360
```



Search completed: August 9, 2004, 19:50:04
Job time : 8310 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 11:55:50 ; Search time 821 Seconds
(without alignments)
11264.710 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2177	100.0	2177	3	AAA75118	Aaa75118 cDNA enco
2	2043.8	93.9	2089	6	AAD29792	Aad29792 Human SCI
3	1913.4	87.9	2413	4	AAS01551	Aas01551 Human sec
4	1893	87.0	2263	4	AAH23100	Aah23100 Osteoarth
5	1808.8	83.1	2589	6	AAD29777	Aad29777 Human SCI
6	1807.8	83.0	2039	6	AAD29791	Aad29791 Human SCI
7	1807.8	83.0	2145	6	AAD29793	Aad29793 Human SCI
8	1794.4	82.4	1959	6	AAL47918	Aal47918 Human cho
9	1794.4	82.4	2507	6	AAL47917	Aal47917 Human cho
10	1794.4	82.4	2507	7	ACC50982	Acc50982 Human bla
11	1794.4	82.4	2507	7	ABX76365	Abx76365 Lung canc
12	1793.4	82.4	1947	6	AAL47936	Aal47936 Human cho
13	1687.8	77.5	1849	6	AAS62301	Aas62301 cDNA sequ
14	1408.8	64.7	2178	4	AAH14000	Aah14000 Human cDN
15	1408.8	64.7	2178	7	ACC50981	Acc50981 Human bla
16	1408.8	64.7	2178	7	ABX76364	Abx76364 Lung canc
17	1349	62.0	1574	4	AAI60925	Aai60925 Human pol
18	1345.8	61.8	1501	4	AAI59139	Aai59139 Human pol
19	1345.8	61.8	1501	8	ADB49122	Adb49122 Novel hum
20	1025.6	47.1	1143	3	AAZ65257	Aaz65257 Human sec
21	1025.6	47.1	1143	7	ADA56306	Ada56306 Gene enco
22	1025.6	47.1	1143	7	ADA40142	Ada40142 Human sec
23	1025.6	47.1	1143	9	ADE11646	Ade11646 Human sec
24	747.8	34.4	883	6	ABL42156	Abl42156 Nucleotid
25	687.2	31.6	4794	7	ACC50983	Acc50983 Human bla
26	687.2	31.6	4794	7	ABX76366	Abx76366 Lung canc
27	556.8	25.6	608	4	AAH05239	Aah05239 Human cDN
28	429	19.7	465	8	ACH31964	Ach31964 Human end
29	397.4	18.3	447	8	ACH42110	Ach42110 Human foe
30	228.4	10.5	319	2	AAT25375	Aat25375 Human gen
c 31	191.2	8.8	1400	5	AAS83195	Aas83195 DNA encod
c 32	101	4.6	1216	6	ABQ28697	Abq28697 Oligonucl
33	101	4.6	1216	6	ABQ28696	Abq28696 Oligonucl
c 34	72.8	3.3	1216	6	ABQ28694	Abq28694 Oligonucl
35	72.8	3.3	1216	6	ABQ28695	Abq28695 Oligonucl
36	61.4	2.8	2000	7	ADA71938	Ada71938 Rice gene
37	60	2.8	60	6	ABN43600	Abn43600 Human spl
c 38	58.2	2.7	110000	4	AAI99683_37	Continuation (38 o
39	57	2.6	57	6	AAL47919	Aal47919 Human cho
c 40	55.6	2.6	2000	7	ADA71938	Ada71938 Rice gene
c 41	54.8	2.5	110000	4	AAI99682_37	Continuation (38 o
42	49.2	2.3	10732	3	AAA10594	Aaa10594 Gene enco
c 43	48.6	2.2	985	6	ABQ44816	Abq44816 Oligonucl
44	48.6	2.2	985	6	ABQ44817	Abq44817 Oligonucl
45	48.2	2.2	1815	5	AAD09400	Aad09400 Zea mays

ALIGNMENTS

RESULT 1
AAA75118
ID AAA75118 standard; cDNA; 2177 BP.
XX
AC AAA75118;
XX
DT 15-JAN-2001 (first entry)
XX
DE cDNA encoding a human secretory protein.
XX
KW Human; secretory protein; HSECP; cancer; gastrointestinal disorder;
KW inflammation; cardiovascular disorder; neurological disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 123. .2003
FT /*tag= a
FT /product= "secretory protein"
FT sig_peptide 123. .203
FT /*tag= b
XX
PN WO200052151-A2.
XX
PD 08-SEP-2000.
XX
PF 03-MAR-2000; 2000WO-US005621.
XX
PR 05-MAR-1999; 99US-0123117P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Baughn MR, Yue H, Au-Young J, Lu DAM, Azimzai Y;
XX
DR WPI; 2000-579282/54.
DR P-PSDB; AAB08864.
XX
PT Twenty two human secretory proteins for diagnosing, treating and
PT preventing cancer, inflammation, and gastrointestinal, cardiovascular and
PT neurological disorders.
XX
PS Claim 4; Page 102-103; 107pp; English.
XX
CC The present sequence encodes a human secretory protein, designated HSECP-
CC 1. The specification also describes HSECP-2 to HSECP-22. The proteins are
CC useful for diagnosing, treating and preventing cancer, inflammation, and

CC gastrointestinal, cardiovascular and neurological disorders. The proteins
CC may also be used to identify agonists, antagonists, and inhibitors. The
CC polynucleotides may be used for producing the protein recombinantly, and
CC as a source of probes and primers for isolating and identifying related
CC sequences
XX
SQ Sequence 2177 BP; 466 A; 627 C; 660 G; 424 T; 0 U; 0 Other;

Query Match 100.0%; Score 2177; DB 3; Length 2177;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60

Qy   61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120

Qy  121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCT 180

Qy  181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCA 240

Qy  241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAG 300

Qy  301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA 360

Qy  361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 ACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGC 420

Qy  421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT 480

Qy  481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCT 540

Qy  541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACA 600

Qy  601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTG 660

Qy  661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCT 720

Qy  721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780

Qy  781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840

Qy  841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACA 900

Qy  901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960

Qy  961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020

Qy 1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080

Qy 1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140

Qy 1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200

Qy 1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCA 1260

Qy 1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320

Qy 1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380

Qy 1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGG 1440

Qy 1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 GCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGG 1500

Qy 1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGG 1560

Qy 1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 GCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAG 1620

Qy 1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 CCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCG 1680

Qy 1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 GGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACC 1740

Qy 1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 CAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACA 1800

Qy 1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 CCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACA 1860

Qy 1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Qy 1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1921 AGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAG 1980

Qy 1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
```

```
Db       1981 TTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATA 2040
Qy       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTA 2100
Qy       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TTAAGCTATTAATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCC 2160
Qy       2161 TCTTGAAAAAAAAAAAA 2177
             |||||||||||||||||
Db       2161 TCTTGAAAAAAAAAAAA 2177
```

RESULT 2
AAD29792
ID AAD29792 standard; cDNA; 2089 BP.
XX
AC AAD29792;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 splice variant B cDNA.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 59. .1915
FT /*tag= a
FT /product= "Human SCIM-1 splice variant B protein"
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR P-PSDB; AAE18681.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 4; Page 36-37; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for
CC induction of immunological tolerance to an autoantigen in patients
CC suffering from autoimmune disorders, more specifically rheumatoid
CC arthritis. The inflammatory disorders treated include diseases like
CC Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC multiple sclerosis and diabetes. Polypeptides of the invention are also
CC useful in a diagnostic method for the detection of activated autoreactive
CC T cells. The present sequence is human SCIM-1 splice variant B cDNA
XX
SQ Sequence 2089 BP; 462 A; 592 C; 614 G; 421 T; 0 U; 0 Other;

Query Match 93.9%; Score 2043.8; DB 6; Length 2089;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2045; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102
Qy        191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162
Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222
Qy        311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282
Qy        371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342
Qy        431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402
Qy        491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462
Qy        551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522
Qy        611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582
Qy        671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642
Qy        731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702
Qy        791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762
Qy        851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822
Qy        911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882
Qy        971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942
Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy       1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy       1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy       1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy       1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy       1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy       1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy       1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy       1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy       1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy       1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy       1931 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1990
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1902

Qy       1991 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 2050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 1962

Qy       2051 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2022

Qy       2111 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2082

Qy       2171 AAAAAAA 2177
              |||||||
Db       2083 AAAAAAA 2089
```

RESULT 3
AAS01551
ID AAS01551 standard; cDNA; 2413 BP.
XX
AC AAS01551;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secretory molecule cDNA sptm #41.
XX
KW Human; secretory molecule; sptm; SPTM; library screening; gene therapy;
KW cell signalling; cell proliferative disorder; atherosclerosis; cancer;
KW immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
KW nervous system disease; mental retardation; developmental disorder;
KW neuromuscular disorder; microarray; Incyte ID number 4814544dec; ss.
XX
OS Homo sapiens.
XX
PN WO200123558-A2.
XX
PD 05-APR-2001.
XX
PF 19-SEP-2000; 2000WO-US025610.
XX
PR 28-SEP-1999; 99US-0156624P.
PR 28-SEP-1999; 99US-0156625P.
PR 02-DEC-1999; 99US-0168611P.
PR 02-DEC-1999; 99US-0168613P.
PR 02-DEC-1999; 99US-0168614P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Shah P, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM;
PI Wright RJ, Chen W, Liu TF, Yap PE, Stockdreher TK, Amshey S;
PI Fong WT;
XX
DR WPI; 2001-258134/26.
XX
PT New secretory polynucleotides (SPTM) and the polypeptides they encode,
PT for use in inducing antibodies and screening libraries of compounds.
XX
PS Claim 1; Page 146; 161pp; English.
XX
CC The present sequence for human secretory molecule cDNA sptm #41 (Incyte
CC ID number 4814544dec) is 1 of 63 novel sptm cDNA sequences (AAS01511-
CC AAS01573) which encode for the secretory polypeptides SPTM. The sptm
CC polynucleotides are useful for screening a compound for effectiveness in
CC altering expression of a target polynucleotide, where the target
CC polynucleotide comprises sptm. Sptm is also useful in a method for
CC assessing the toxicity of a test compound. Sptm and its fragments or
CC complementary sequences, may be used to identify the presence of and/or
CC determine the degree of similarity between two nucleic acid sequences.
CC Sptm can also be used for a variety of diagnostic and therapeutic
CC purposes, e.g. diagnosing a particular condition, disease or disorder
CC associated with cell signalling, such conditions include cell
CC proliferative disorders such as atherosclerosis, and cancers including
CC leukaemia, an immune system disorder e.g. acquired immunodeficiency
CC syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
CC disease, nutritional and metabolic disease of the nervous system, mental
CC retardation and other developmental disorders, and muscular dystrophy and
CC other neuromuscular disorders. Sptm can also be used to design probes
CC useful in diagnostic assays, which may be used to monitor the progress of
CC conditions or disorders associated with abnormal levels of expression of
CC sptm. In addition sptm encoding SPTM may be used for somatic or germline
CC gene therapy, for inducing antibodies, or in microarrays

```
XX
SQ   Sequence 2413 BP; 511 A; 714 C; 722 G; 459 T; 0 U; 7 Other;

  Query Match           87.9%;  Score 1913.4;  DB 4;  Length 2413;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 1917;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
Db        318 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 377

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
Db        378 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 437

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
Db        438 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 497

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
Db        498 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 557

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
Db        558 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 617

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
Db        618 ATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 677

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
Db        678 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 737

Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
Db        738 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 797

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
Db        798 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 857

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
Db        858 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 917

Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
Db        918 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGATGCTCTGTGGCCTGTGTGGACAGAA 977

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
Db        978 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 1037

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 798
Db       1038 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA 1097

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
Db       1098 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 1157

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
Db       1158 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1217

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
Db       1218 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1277

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
Db       1278 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1337

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
Db       1338 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1397

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
Db       1398 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1457

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
Db       1458 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1517

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
Db       1518 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1577

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
Db       1578 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1637

Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
Db       1638 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1697

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
Db       1698 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1757

Qy       1459 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1518
Db       1758 TGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCT 1817

Qy       1519 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1578
Db       1818 ACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTG 1877

Qy       1579 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1638
Db       1878 AGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGA 1937

Qy       1639 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1698
Db       1938 CGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGC 1997

Qy       1699 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 1758
Db       1998 TGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTG 2057

Qy       1759 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 1818
Db       2058 GCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGT 2117

Qy       1819 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 1878
Db       2118 TCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGT 2177

Qy       1879 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 1938
Db       2178 GCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG 2237

Qy       1939 TGG 1941
Db       2238 TGG 2240


RESULT 4
AAH23100
ID   AAH23100 standard; DNA; 2263 BP.
XX
AC   AAH23100;
XX
```

DT 17-SEP-2001 (first entry)
XX
DE Osteoarthritis tissue-derived nucleic acid sequence #30.
XX
KW Osteoarthritis; infectious disorder; Crohn's disease; sepsis; human;
KW wound healing; osteopathic; anti-arthritic, anti-inflammatory; vulnerary;
KW antibacterial; antiallergic; ds.
XX
OS Homo sapiens.
XX
PN WO200153531-A2.
XX
PD 26-JUL-2001.
XX
PF 18-JAN-2001; 2001WO-US000016.
XX
PR 18-JAN-2000; 2000US-0176523P.
XX
PA (PHAA) PHARMACIA CORP.
XX
PI Phippard D, Vasanthakamur G, Dotson S, Ma X;
XX
DR WPI; 2001-451914/48.
XX
PT Substantially purified protein, polypeptide or their fragments, used to
PT identify a biologically active compound or composition and treat
PT mammalian osteoarthritis.
XX
PS Claim 1; Page 115-116; 144pp; English.
XX
CC Sequences AAH23071-23152 represent nucleic acid sequences derived from
CC osteoarthritis tissues. The sequences are useful as probes and for the
CC diagnosis or prognosis of mammalian osteoarthritis. The polynucleotides
CC and polypeptides of the invention are useful for generating diagnostic
CC reagents, as targets for small molecule drug development, generation of
CC therapeutics, and cloning genes. Specific antibodies are used to generate
CC enzyme linked immunosorbent assays for detection of osteoarthritis. The
CC invented molecules can be used to treat osteoarthritis or to analyse the
CC disease-modifying activity of osteoarthritis drugs. Other disorders
CC treatable using the nucleic acid sequences include atopic, inflammatory
CC and infectious disorders e.g. Crohn's disease and sepsis, and wound
CC healing
XX
SQ Sequence 2263 BP; 485 A; 667 C; 676 G; 428 T; 0 U; 7 Other;

Query Match 87.0%; Score 1893; DB 4; Length 2263;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1918; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```
Qy        19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC  78
Db       153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC  212

Qy        79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG  138
Db       213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG  272

Qy       139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG  198
Db       273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG  332

Qy       199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG  258
Db       333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG  392

Qy       259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG  318
Db       393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG  452

Qy       319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG  378
Db       453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG  512

Qy       379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG  438
Db       513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG  572

Qy       439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG  498
Db       573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG  632

Qy       499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA  558
Db       633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA  692

Qy       559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA  618
Db       693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA  752

Qy       619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA  678
Db       753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA  812

Qy       679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG  738
Db       813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG  872

Qy       739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA  798
Db       873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA  932

Qy       799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG  858
Db       933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG  992

Qy       859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT  918
Db       993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT  1052

Qy       919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG  978
Db      1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG  1112

Qy       979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG  1038
Db      1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG  1172

Qy      1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC  1098
Db      1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC  1232

Qy      1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC  1158
Db      1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC  1292

Qy      1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA  1218
Db      1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA  1352

Qy      1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG  1278
Db      1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG  1412

Qy      1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG  1338
Db      1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG  1472

Qy      1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG  1398
Db      1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG  1532

Qy      1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC  1458
Db      1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC  1592

Qy      1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC  1517
```

```
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652

Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577

Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712

Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637

Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772

Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697

Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832

Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1757

Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1892

Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816

Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952

Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876

Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012

Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936

Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072

Qy       1937 CGTGG 1941

Db       2073 CGTGG 2077
```

RESULT 5
AAD29777
ID AAD29777 standard; cDNA; 2589 BP.
XX
AC AAD29777;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 protein encoding cDNA.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   59. .2020
FT   /*tag= a
FT   /product= "Human SCIM-1 protein"
FT sig_peptide   59. .121
FT   /*tag= d
FT sig_peptide   59. .115
FT   /*tag= b
FT mat_peptide   116. .2017
FT   /*tag= c
FT   /product= "Human SCIM-1 mature protein #1"
FT mat_peptide   122. .2017
FT   /*tag= e
FT   /product= "Human SCIM-1 mature protein #2"
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR P-PSDB; AAE18679.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 4; Page 29-30; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for
CC induction of immunological tolerance to an autoantigen in patients
CC suffering from autoimmune disorders, more specifically rheumatoid
CC arthritis. The inflammatory disorders treated include diseases like
CC Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC multiple sclerosis and diabetes. Polypeptides of the invention are also
CC useful in a diagnostic method for the detection of activated autoreactive
CC T cells. The present sequence is human SCIM-1 protein encoding cDNA
XX
SQ Sequence 2589 BP; 585 A; 726 C; 765 G; 512 T; 0 U; 1 Other;

```
  Query Match            83.1%;  Score 1808.8;  DB 6;  Length 2589;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1810;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190

Db         43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy        191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250

Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310

Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy        311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370

Db        223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy        371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430

Db        283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy        431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490

Db        343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy        491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550

Db        403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy        551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610

Db        463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522
```

```
Qy       611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db       523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy       671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db       583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy       731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
Db       643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy       791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db       703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy       851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db       763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy       911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db       823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy       971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db       883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy      1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db       943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy      1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy      1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy      1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy      1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy      1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy      1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy      1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy      1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy      1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy      1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy      1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy      1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy      1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy      1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy      1931 AGCCTGCGTGGG 1942
Db       1843 AGCCTGCGTGGG 1854
```

RESULT 6
AAD29791
ID AAD29791 standard; cDNA; 2039 BP.
XX
AC AAD29791;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human SCIM-1 splice variant A cDNA.
XX
KW Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW extracellular matrix protein; tolerogenic protein; Grave's disease;
KW autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   59. .1972
FT   /*tag= a
FT   /product= "Human SCIM-1 splice variant A protein"
XX
PN WO200206478-A2.
XX
PD 24-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-EP007888.
XX
PR 13-JUL-2000; 2000EP-00202495.
XX
PA (ALKU) AKZO NOVEL NV.
XX
PI Heus HC, Nelissen RLH, Meeuwisse CML;
XX
DR WPI; 2002-179796/23.
DR P-PSDB; AAE18680.
XX
PT New extracellular matrix protein useful for prevention of inflammatory
PT diseases, more specifically in induction of T cell tolerance to the
PT protein in patients suffering from rheumatoid arthritis.
XX
PS Claim 4; Page 35-36; 45pp; English.
XX
CC The invention relates to human extracellular matrix protein i.e
CC Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC acid molecules encoding such proteins. Sequences of the invention are
CC useful in a screening assay for the identification of tolerogenic
CC polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC a pharmaceutical preparation against inflammatory diseases and for

```
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 splice variant A cDNA
XX
SQ   Sequence 2039 BP; 448 A; 586 C; 596 G; 409 T; 0 U; 0 Other;

  Query Match             83.0%;  Score 1807.8;  DB 6;  Length 2039;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
Db        43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy       191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
Db       103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy       251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
Db       163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy       311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
Db       223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy       371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
Db       283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy       431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
Db       343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy       491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
Db       403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy       551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
Db       463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy       611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db       523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy       671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db       583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy       731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
Db       643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy       791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db       703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy       851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db       763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy       911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db       823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy       971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db       883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy      1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db       943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy      1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy      1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy      1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy      1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy      1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy      1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy      1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy      1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy      1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy      1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy      1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy      1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy      1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy      1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy      1931 AGCCTGCGTGG 1941
Db       1843 AGCCTGCGTGG 1853


RESULT 7
AAD29793
ID   AAD29793 standard; cDNA; 2145 BP.
XX
```

```
AC   AAD29793;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human SCIM-1 splice variant C cDNA.
XX
KW   Human; synovium/cartilage inflammation-linked messenger-1; SCIM-1;
KW   extracellular matrix protein; tolerogenic protein; Grave's disease;
KW   autoimmune disorder; juvenile arthritis; primary glomerulonephritis;
KW   polyarthritis; osteoarthritis; Sjogren's syndrome; myasthenia gravis;
KW   rheumatoid arthritis; Addison's disease; primary biliary sclerosis;
KW   uveitis; systemic lupus erythematosus; inflammatory bowel disease;
KW   multiple sclerosis; diabetes; inflammatory disorder; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             59. .1897
FT                   /*tag= a
FT                   /product= "Human SCIM-1 splice variant C protein"
XX
PN   WO200206478-A2.
XX
PD   24-JAN-2002.
XX
PF   09-JUL-2001; 2001WO-EP007888.
XX
PR   13-JUL-2000; 2000EP-00202495.
XX
PA   (ALKU ) AKZO NOVEL NV.
XX
PI   Heus HC,  Nelissen RLH,  Meeuwisse CML;
XX
DR   WPI; 2002-179796/23.
DR   P-PSDB; AAE18682.
XX
PT   New extracellular matrix protein useful for prevention of inflammatory
PT   diseases, more specifically in induction of T cell tolerance to the
PT   protein in patients suffering from rheumatoid arthritis.
XX
PS   Claim 4; Page 37; 45pp; English.
XX
CC   The invention relates to human extracellular matrix protein i.e
CC   Synovium/Cartilage Inflammation-linked Messenger-1 (SCIM-1) and nucleic
CC   acid molecules encoding such proteins. Sequences of the invention are
CC   useful in a screening assay for the identification of tolerogenic
CC   polypeptides. SCIM proteins are useful in therapy and for manufacture of
CC   a pharmaceutical preparation against inflammatory diseases and for
CC   induction of immunological tolerance to an autoantigen in patients
CC   suffering from autoimmune disorders, more specifically rheumatoid
CC   arthritis. The inflammatory disorders treated include diseases like
CC   Grave's diseases, juvenile arthritis, primary glomerulonephritis,
CC   polyarthritis, osteoarthritis, Sjogren's syndrome, myasthenia gravis,
CC   rheumatoid arthritis, Addison's disease, primary biliary sclerosis,
CC   uveitis, systemic lupus erythematosus, inflammatory bowel disease,
CC   multiple sclerosis and diabetes. Polypeptides of the invention are also
CC   useful in a diagnostic method for the detection of activated autoreactive
CC   T cells. The present sequence is human SCIM-1 splice variant C cDNA
XX
SQ   Sequence 2145 BP; 497 A; 604 C; 611 G; 433 T; 0 U; 0 Other;

  Query Match             83.0%;  Score 1807.8;  DB 6;  Length 2145;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy       191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy       251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy       311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy       371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy       431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy       491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy       551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy       611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy       671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy       731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy       791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy       851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy       911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy       971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy      1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy      1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy      1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy      1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy      1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy      1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
```

```
Db       1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450

Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510

Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy       1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570

Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy       1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630

Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy       1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690

Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy       1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750

Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy       1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810

Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy       1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870

Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy       1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930

Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy       1931 AGCCTGCGTGG 1941

Db       1843 AGCCTGCGTGG 1853
```

```
RESULT 8
AAL47918
ID   AAL47918 standard; cDNA; 1959 BP.
XX
AC   AAL47918;
XX
DT   26-SEP-2002  (first entry)
XX
DE   Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:2.
XX
KW   Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW   extracellular matrix protein; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .1959
FT                   /*tag=  a
FT                   /product= "CEP-68"
FT                   /partial
FT                   /note= "no stop codon"
FT   sig_peptide     1. .57
FT                   /*tag=  b
FT   mat_peptide     58. .1959
FT                   /*tag=  c
XX
PN   WO200253709-A1.
XX
PD   11-JUL-2002.
XX
PF   24-DEC-2001; 2001WO-EP015307.
XX
PR   05-JAN-2001; 2001DE-01000305.
XX
PA   (CYTO-) CYTONET GMBH & CO KG.
XX
PI   Richter W,  Steck E;
XX
DR   WPI; 2002-528856/56.
DR   P-PSDB; AAO18264.
XX
PT   New nucleic acid encoding an extracellular matrix protein, useful, for
PT   example, as a marker for detecting or isolating chondrocytes.
XX
PS   Claim 2; Page 92; 107pp; German.
XX
CC   The present invention relates to the protein and coding sequences of
CC   novel human extracellular matrix proteins, more specifically chondrocyte
CC   specific CEP-68 proteins. The sequences can be used in the identification
CC   of cell differentiation stages in cells which develop from mesenchymal
CC   stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC   present sequence is a coding sequence of the CEP-68 protein
XX
SQ   Sequence 1959 BP; 394 A; 590 C; 596 G; 379 T; 0 U; 0 Other;
```

```
  Query Match             82.4%;  Score 1794.4;  DB 6;  Length 1959;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206

Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266

Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326

Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386

Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446

Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506

Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566

Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626

Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686

Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746

Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806

Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660
```

```
Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740
```

```
Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 9
AAL47917
ID AAL47917 standard; cDNA; 2507 BP.
XX
AC AAL47917;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:1.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein; gene; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .1962
FT		/*tag= a
FT		/product= "CEP-68"
FT	sig_peptide	1. .57
FT		/*tag= b
FT	mat_peptide	58. .1959
FT		/*tag= c

XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR P-PSDB; AAO18264.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 1; Page 91-92; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is a coding sequence of the CEP-68 protein
XX
SQ Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

Query Match 82.4%; Score 1794.4; DB 6; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180
```

```
Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 10
ACC50982
ID ACC50982 standard; cDNA; 2507 BP.
XX
AC ACC50982;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human bladder cancer associated cDNA sequence SEQ ID NO:57.
XX
KW Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003003906-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US021338.
XX
PR 03-JUL-2001; 2001US-0302814P.
PR 03-AUG-2001; 2001US-0310099P.
PR 08-NOV-2001; 2001US-0343705P.
PR 13-NOV-2001; 2001US-0350666P.
PR 12-APR-2002; 2002US-0372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Mack DH, Aziz N;
XX
DR WPI; 2003-201532/19.
DR P-PSDB; ABR48171.
XX
PT Detecting a bladder cancer-associated transcript in a cell from a
PT patient, comprises contacting a biological sample from the patient with a
PT bladder cancer-associated polynucleotide or antibody.

XX
PS Claim 6; Page 248-249; 307pp; English.
XX
CC The present invention describes a method for detecting a bladder cancer-
CC associated transcript in a cell from a patient. The method comprises
CC contacting a biological sample from the patient with a polynucleotide
CC that selectively hybridises to a sequence that is 80 % identical to a
CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC encode the human bladder cancer-associated proteins given in ABR48146 to
CC ABR48242). Bladder cancer-associated sequences from the present invention
CC have cytostatic activities, and can be used in antisense gene therapy and
CC in vaccine production. The method can be used for detecting a bladder
CC cancer-associated transcript in a cell from a patient. The method is
CC useful in diagnosing or treating bladder cancer and in screening for
CC compounds that modulate bladder cancer, such as hormones or antibodies.
CC The nucleic acid molecules from the present invention may be used in
CC various screening and diagnostic methods, and for gene therapy, vaccine
CC and/or antisense/inhibition applications
XX
SQ Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

Query Match 82.4%; Score 1794.4; DB 7; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy      1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy      1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy      1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy      1707 CTCTACCCCCGGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy      1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy      1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy      1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 11
ABX76365
ID ABX76365 standard; DNA; 2507 BP.
XX
AC ABX76365;
XX
DT 02-APR-2003 (first entry)
XX
DE Lung cancer-associated polynucleotide #229.
XX
KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS Unidentified.
XX
PN WO200286443-A2.
XX
PD 31-OCT-2002.
XX
PF 18-APR-2002; 2002WO-US012476.
XX
PR 18-APR-2001; 2001US-0284770P.
PR 10-MAY-2001; 2001US-0290492P.
PR 09-NOV-2001; 2001US-0339245P.
PR 13-NOV-2001; 2001US-0350666P.
PR 29-NOV-2001; 2001US-0334370P.
PR 12-APR-2002; 2002US-0372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Aziz N, Murray R;
XX
DR WPI; 2003-093161/08.
DR P-PSDB; ABU56636.
XX
PT Detecting a lung cancer-associated transcript in a cell from a patient
PT for treating lung cancer, by contacting a biological sample from the
PT patient with a polynucleotide that exhibits increased or decreased
PT expression in lung cancer.
XX
PS Claim 22; Page 366; 453pp; English.
XX
CC The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated polypeptide, for
CC inhibiting proliferation of a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer, such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
CC invention
XX
SQ Sequence 2507 BP; 546 A; 710 C; 746 G; 505 T; 0 U; 0 Other;

Query Match 82.4%; Score 1794.4; DB 7; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140
```

```
Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 12
AAL47936
ID AAL47936 standard; cDNA; 1947 BP.
XX
AC AAL47936;
XX
DT 26-SEP-2002 (first entry)
XX
DE Human chondrocyte specific CEP-68 protein coding sequence SEQ ID NO:22.
XX
KW Human; chondrocyte specific protein; CEP-68; cartilage; marker;
KW extracellular matrix protein; gene; ss.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	1. .1914
FT		/*tag= a
FT		/product= "alternatively spliced CEP-68"
FT		/transl_except= (pos:1129. .1131,aa:Glu)
FT		/transl_except= (pos:1297. .1299,aa:Phe)
FT		/transl_except= (pos:1621. .1623,aa:Val)
FT		/transl_except= (pos:1648. .1650,aa:Asn)
FT		/transl_except= (pos:1669. .1674,aa:Phe)

XX
PN WO200253709-A1.
XX
PD 11-JUL-2002.
XX
PF 24-DEC-2001; 2001WO-EP015307.
XX
PR 05-JAN-2001; 2001DE-01000305.
XX
PA (CYTO-) CYTONET GMBH & CO KG.
XX
PI Richter W, Steck E;
XX
DR WPI; 2002-528856/56.
DR P-PSDB; AAO18267.
XX
PT New nucleic acid encoding an extracellular matrix protein, useful, for
PT example, as a marker for detecting or isolating chondrocytes.
XX
PS Claim 4; Page 97-98; 107pp; German.
XX
CC The present invention relates to the protein and coding sequences of
CC novel human extracellular matrix proteins, more specifically chondrocyte
CC specific CEP-68 proteins. The sequences can be used in the identification
CC of cell differentiation stages in cells which develop from mesenchymal
CC stem cells, particularly chondrocytes and to isolate chondrocytes. The
CC present sequence is an alternatively spliced coding sequence of the CEP-
CC 68 protein
XX
SQ Sequence 1947 BP; 413 A; 566 C; 571 G; 397 T; 0 U; 0 Other;

Query Match 82.4%; Score 1793.4; DB 6; Length 1947;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
```

```
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660
Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720
Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780
Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840
Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900
Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db        901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960
Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020
Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080
Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140
Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200
Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260
Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320
Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380
Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440
Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500
Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560
Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620
Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680
Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740
Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1795
```

RESULT 13
AAS62301
ID AAS62301 standard; cDNA; 1849 BP.
XX
AC AAS62301;
XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #88 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
XX
OS Homo sapiens.
XX
PN WO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US010485.
XX
PR 06-APR-2000; 2000US-0195604P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-010900/01.
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease.
XX
PS Claim 1; Page 122; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides a
CC method for producing proteins from these polynucleotide sequences. The
CC proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are useful
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),
CC immune deficiency disorders (e.g. severe combined immunodeficiency
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
CC (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of
CC the invention are also useful in gene therapy. AAS62214-AAS62838
CC represent the cDNA sequences of the invention that encode for novel human
CC secreted proteins
XX
SQ Sequence 1849 BP; 395 A; 535 C; 544 G; 372 T; 0 U; 3 Other;

Query Match 77.5%; Score 1687.8; DB 6; Length 1849;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1700; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
Db          1 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 60
Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACC 358
```

```
Db        61 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 120
Qy       359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
Db       121 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 180
Qy       419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
Db       181 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGC 240
Qy       479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
Db       241 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 300
Qy       539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
Db       301 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 360
Qy       599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
Db       361 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 420
Qy       659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 718
Db       421 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 480
Qy       719 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 778
Db       481 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 540
Qy       779 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 838
Db       541 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 600
Qy       839 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898
Db       601 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 660
Qy       899 CAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 958
Db       661 CAATGAGAATGGGCCTAAATTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 720
Qy       959 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 1018
Db       721 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 780
Qy      1019 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 1078
Db       781 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 840
Qy      1079 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 1138
Db       841 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 900
Qy      1139 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 1198
Db       901 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 960
Qy      1199 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1258
Db       961 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1020
Qy      1259 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1318
Db      1021 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1080
Qy      1319 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1378
Db      1081 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1140
Qy      1379 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1438
Db      1141 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1200
Qy      1439 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1498
Db      1201 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1260
Qy      1499 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1558
Db      1261 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1320
Qy      1559 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1618
Db      1321 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1380
Qy      1619 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
Db      1381 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1440
Qy      1679 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1738
Db      1441 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1500
Qy      1739 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1798
Db      1501 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1560
Qy      1799 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1858
Db      1561 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1620
Qy      1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAA 1918
Db      1621 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTC-GGGCTACGAGCCCAA 1679
Qy      1919 CGAGGATGGCACAGCCTGCGTGG 1941
Db      1680 CGAGGATGGCACAGCCTGCGTGG 1702
```

RESULT 14
AAH14000
ID AAH14000 standard; cDNA; 2178 BP.
XX
AC AAH14000;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11082.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 11082; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2178 BP; 477 A; 618 C; 655 G; 428 T; 0 U; 0 Other;

Query Match 64.7%; Score 1408.8; DB 4; Length 2178;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

```
Qy       435 TACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGG 494
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         1 TACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGG 60

Qy       495 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 554
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 120

Qy       555 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 614
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 180

Qy       615 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 674
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 240

Qy       675 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 734
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 300

Qy       735 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 794
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 360

Qy       795 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 854
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 420

Qy       855 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 914
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 480

Qy       915 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 540

Qy       975 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 1034
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 600

Qy      1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 1094
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 660

Qy      1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 1154
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 720

Qy      1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 780

Qy      1215 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 1274
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 840

Qy      1275 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 1334
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 900

Qy      1335 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 1394
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 960

Qy      1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1020

Qy      1455 TGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1514
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1021 TGGCTGCGAGTGGTGCCACGCACCCGGGTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1080

Qy      1515 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1574
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1140

Qy      1575 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200

Qy      1635 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1694
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1260

Qy      1695 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAG 1754
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGA- 1319

Qy      1755 TGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATC 1814
                                                    ||||||||||||||||||
Db       1320 ----------------------------------------GACACCAATGAATGCATC 1337

Qy      1815 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1874
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1397

Qy      1875 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1934
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1457

Qy      1935 TGCGTGGG 1942
            ||||||||
Db       1458 TGCGTGGG 1465
```

RESULT 15
ACC50981
ID ACC50981 standard; cDNA; 2178 BP.
XX
AC ACC50981;
XX
DT 12-JUN-2003 (first entry)

XX
DE Human bladder cancer associated cDNA sequence SEQ ID NO:55.
XX
KW Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003003906-A2.
XX
PD 16-JAN-2003.
XX
PF 03-JUL-2002; 2002WO-US021338.
XX
PR 03-JUL-2001; 2001US-0302814P.
PR 03-AUG-2001; 2001US-0310099P.
PR 08-NOV-2001; 2001US-0343705P.
PR 13-NOV-2001; 2001US-0350666P.
PR 12-APR-2002; 2002US-0372246P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
PI Mack DH, Aziz N;
XX
DR WPI; 2003-201532/19.
DR P-PSDB; ABR48170.
XX
PT Detecting a bladder cancer-associated transcript in a cell from a
PT patient, comprises contacting a biological sample from the patient with a
PT bladder cancer-associated polynucleotide or antibody.
XX
PS Claim 6; Page 247-248; 307pp; English.
XX
CC The present invention describes a method for detecting a bladder cancer-
CC associated transcript in a cell from a patient. The method comprises
CC contacting a biological sample from the patient with a polynucleotide
CC that selectively hybridises to a sequence that is 80 % identical to a
CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059
CC encode the human bladder cancer-associated proteins given in ABR48146 to
CC ABR48242). Bladder cancer-associated sequences from the present invention
CC have cytostatic activities, and can be used in antisense gene therapy and
CC in vaccine production. The method can be used for detecting a bladder
CC cancer-associated transcript in a cell from a patient. The method is
CC useful in diagnosing or treating bladder cancer and in screening for
CC compounds that modulate bladder cancer, such as hormones or antibodies.
CC The nucleic acid molecules from the present invention may be used in
CC various screening and diagnostic methods, and for gene therapy, vaccine
CC and/or antisense/inhibition applications
XX
SQ Sequence 2178 BP; 477 A; 618 C; 655 G; 428 T; 0 U; 0 Other;

Query Match 64.7%; Score 1408.8; DB 7; Length 2178;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

```
Qy        435 TACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGG 494
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 TACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGG 60

Qy        495 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 120

Qy        555 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 180

Qy        615 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 240

Qy        675 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 300

Qy        735 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCG 794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCG 360

Qy        795 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 420

Qy        855 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 480

Qy        915 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 540

Qy        975 GTGGACGACCCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGGACGACCCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 600

Qy       1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 660

Qy       1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 720

Qy       1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 780

Qy       1215 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 840

Qy       1275 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 1334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 900

Qy       1335 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 1394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 960

Qy       1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1020

Qy       1455 TGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1514
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1021 TGGCTGCGAGTGGTGCCACGCACCCGGGTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1080

Qy       1515 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1140

Qy       1575 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200

Qy       1635 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1260

Qy       1695 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAG 1754
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGA- 1319

Qy       1755 TGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATC 1814
                                                   ||||||||||||||||||
Db       1320 ----------------------------------------GACACCAATGAATGCATC 1337
```

```
Qy       1815 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1397

Qy       1875 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1457

Qy       1935 TGCGTGGG 1942
              ||||||||
Db       1458 TGCGTGGG 1465
```

Search completed: August 9, 2004, 17:31:02
Job time : 840 secs

Page Blank (uspto)

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 17:17:00 ; Search time 179 Seconds
(without alignments)
6749.319 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1345.8	61.8	1501	4	US-09-620-312D-1032	Sequence 1032, Ap
	2	788	36.2	789	4	US-09-023-655-484	Sequence 484, App
	3	109	5.0	707	4	US-09-023-655-209	Sequence 209, App
c	4	58.2	2.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	5	54.8	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	6	47.8	2.2	1221	4	US-09-252-991A-10929	Sequence 10929, A
	7	47.8	2.2	1269	4	US-09-252-991A-11061	Sequence 11061, A
c	8	47.8	2.2	1641	4	US-09-252-991A-11134	Sequence 11134, A
	9	46.8	2.1	501	4	US-09-252-991A-10991	Sequence 10991, A
	10	43.8	2.0	430	4	US-09-621-976-16656	Sequence 16656, A
	11	43.8	2.0	1176	4	US-09-252-991A-547	Sequence 547, App
	12	43.8	2.0	1983	4	US-09-252-991A-531	Sequence 531, App
	13	43.4	2.0	1596	4	US-09-252-991A-5938	Sequence 5938, Ap
c	14	43.4	2.0	1674	4	US-09-252-991A-5894	Sequence 5894, Ap
c	15	43	2.0	486	4	US-09-489-039A-2746	Sequence 2746, Ap
	16	43	2.0	798	4	US-09-489-039A-2833	Sequence 2833, Ap
	17	43	2.0	1575	4	US-09-266-965-84	Sequence 84, Appl
	18	43	2.0	18034	4	US-09-266-965-75	Sequence 75, Appl
	19	42.8	2.0	1971	4	US-09-252-991A-8432	Sequence 8432, Ap
	20	42.8	2.0	2535	4	US-09-252-991A-8466	Sequence 8466, Ap
c	21	42.8	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	22	42.6	2.0	2469	4	US-09-252-991A-5349	Sequence 5349, Ap
c	23	42.6	2.0	5103	4	US-09-252-991A-5192	Sequence 5192, Ap
	24	42.6	2.0	6876	4	US-09-252-991A-5283	Sequence 5283, Ap
	25	42.2	1.9	1027	4	US-09-674-741-9	Sequence 9, Appli
	26	42	1.9	687	4	US-09-252-991A-2644	Sequence 2644, Ap
c	27	42	1.9	804	4	US-09-252-991A-2187	Sequence 2187, Ap
c	28	42	1.9	1482	4	US-09-252-991A-2267	Sequence 2267, Ap
	29	42	1.9	2190	4	US-09-252-991A-10256	Sequence 10256, A
	30	42	1.9	2466	4	US-09-266-965-83	Sequence 83, Appl
c	31	42	1.9	2955	4	US-09-252-991A-684	Sequence 684, App
	32	42	1.9	3150	4	US-09-252-991A-727	Sequence 727, App
	33	41.8	1.9	364	4	US-09-621-976-17202	Sequence 17202, A
c	34	41.8	1.9	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	35	41.8	1.9	28958	1	US-08-456-837-6	Sequence 6, Appli
c	36	41.8	1.9	28958	1	US-08-457-342-6	Sequence 6, Appli
c	37	41.8	1.9	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	38	41.8	1.9	28958	1	US-08-458-076A-6	Sequence 6, Appli
c	39	41.8	1.9	28958	1	US-08-764-233A-4	Sequence 4, Appli
c	40	41.8	1.9	28958	1	US-08-457-335A-6	Sequence 6, Appli
c	41	41.8	1.9	28958	1	US-08-729-214-6	Sequence 6, Appli
c	42	41.8	1.9	28958	3	US-09-028-934-6	Sequence 6, Appli
c	43	41.8	1.9	49377	1	US-08-764-233A-1	Sequence 1, Appli
	44	41.6	1.9	942	4	US-09-252-991A-14866	Sequence 14866, A
	45	41.6	1.9	1170	4	US-09-252-991A-5515	Sequence 5515, Ap

ALIGNMENTS

RESULT 1
US-09-620-312D-1032
; Sequence 1032, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1032
; LENGTH: 1501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(1468)
US-09-620-312D-1032

Query Match 61.8%; Score 1345.8; DB 4; Length 1501;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        593 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGG 652
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db          1 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCTTGTGGTGTGGCCAGCCTCTTTGCCGG 60
```

```
Qy        653 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 120

Qy        713 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 180

Qy        773 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 240

Qy        833 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 300

Qy        893 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 360

Qy        953 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 420

Qy       1013 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 480

Qy       1073 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 540

Qy       1133 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 600

Qy       1193 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 660

Qy       1253 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 720

Qy       1313 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 780

Qy       1373 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 1432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 840

Qy       1433 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 1492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 900

Qy       1493 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 1552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 960

Qy       1553 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1612
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1020

Qy       1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1672
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1080

Qy       1673 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1140

Qy       1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1792
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1200

Qy       1793 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1852
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1260

Qy       1853 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1912
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1320

Qy       1913 GCCCAACGAGGATGGCACAGCCTGCGTGG 1941
              |||||||||||||||||||||||||||||
Db       1321 GCCCAACGAGGATGGCACAGCCTGCGTGG 1349
```

RESULT 2
US-09-023-655-484
; Sequence 484, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 484:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 192279
US-09-023-655-484

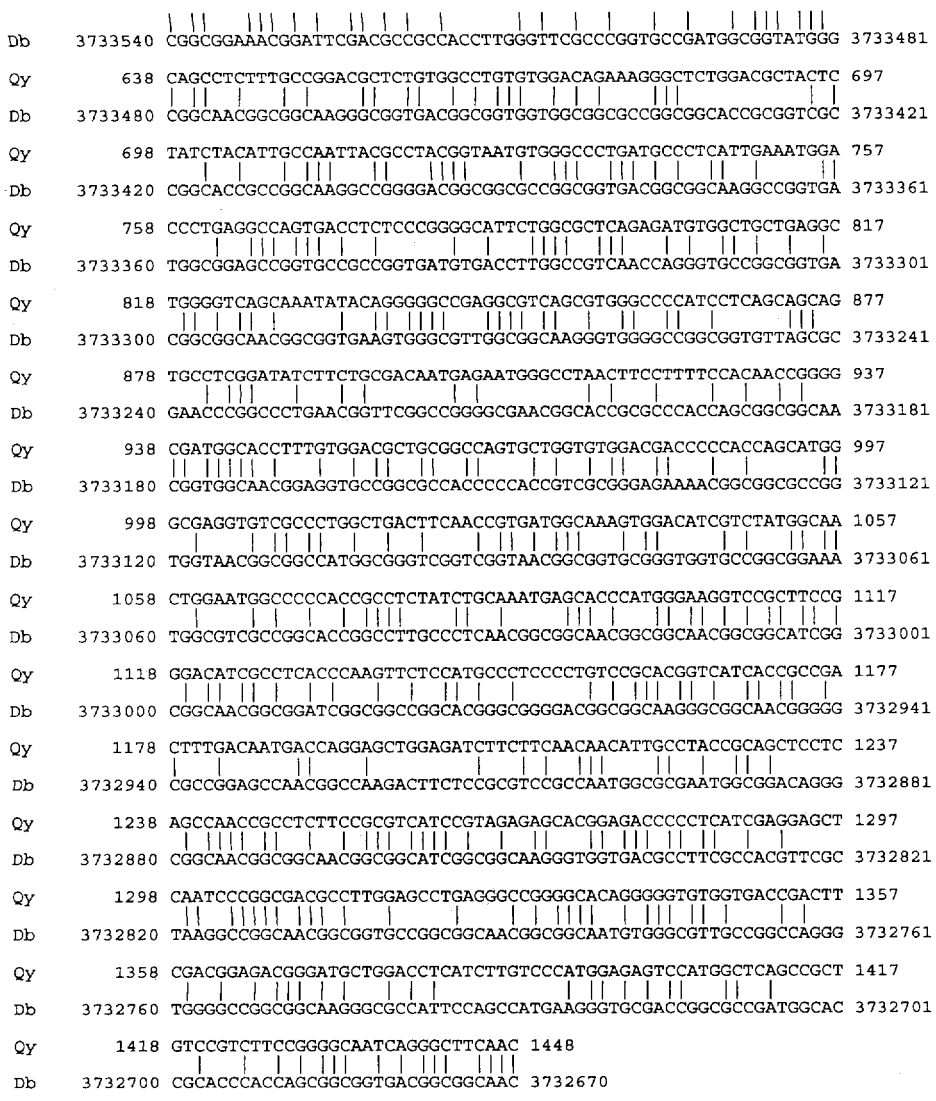
Query Match 36.2%; Score 788; DB 4; Length 789;
Best Local Similarity 99.9%; Pred. No. 3.8e-200;
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1058 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 1117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 60
```

```
Qy       1118 GGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 1177
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGACATCGNCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 120

Qy       1178 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 180

Qy       1238 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 240

Qy       1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 300

Qy       1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 360

Qy       1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 420

Qy       1478 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 480

Qy       1538 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 540

Qy       1598 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 1657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 600

Qy       1658 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 660

Qy       1718 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 1777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 720

Qy       1778 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 1837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 780

Qy       1838 AGACAAGCC 1846
             |||||||||
Db         781 AGACAAGCC 789
```

```
RESULT 3
US-09-023-655-209
; Sequence 209, Application US/09023655
; Patent No. 6607879
;  GENERAL INFORMATION:
;    APPLICANT:  Cocks, Benjamin G.
;    APPLICANT:  Susan G. Stuart
;    APPLICANT:  Jeffrey J. Seilhamer
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
;    TITLE OF INVENTION:  EXPRESSION
;    NUMBER OF SEQUENCES:  1508
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,655
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SYNORAB01
;      CLONE:  079378
US-09-023-655-209

  Query Match             5.0%;  Score 109;  DB 4;  Length 707;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-19;
  Matches  109;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 1893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 60

Qy       1894 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 109
```

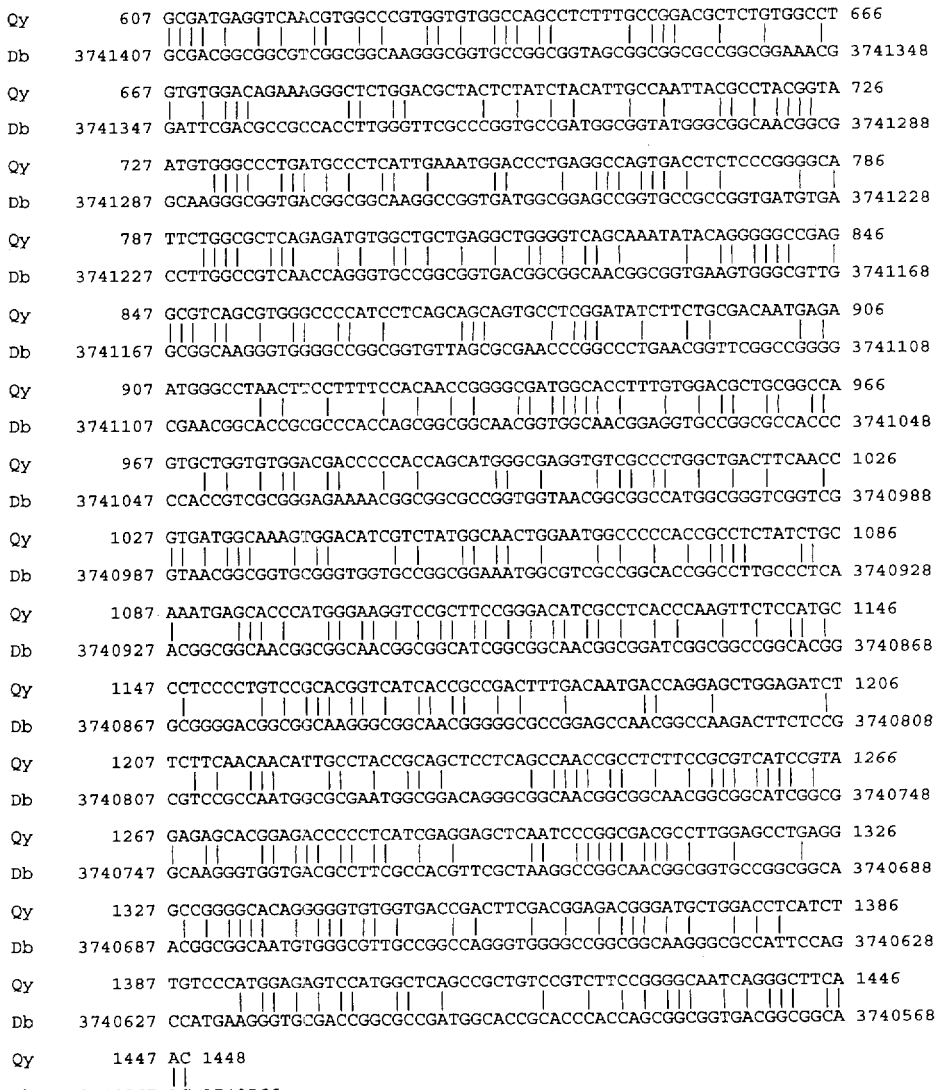
```
RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             2.7%;  Score 58.2;  DB 3;  Length 4403765;
  Best Local Similarity   41.7%;  Pred. No. 0.0012;
  Matches  363;  Conservative    0;  Mismatches  508;  Indels    0;  Gaps    0;

Qy        578 CCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGC 637
```



```
Db    3733540 CGGCGGAAACGGATTCGACGCCGCCACCTTGGGTTCGCCCGGTGCCGATGGCGGTATGGG 3733481

Qy        638 CAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTC 697

Db    3733480 CGGCAACGGCGGCAAGGGCGGTGACGGCGGTGGTGGCGGCGCCGGCGGCACCGCGGTCGC 3733421

Qy        698 TATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGA 757

Db    3733420 CGGCACCGCCGGCAAGGCCGGGGACGGCGGCGCCGGCGGTGACGGCGGCAAGGCCGGTGA 3733361

Qy        758 CCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGC 817

Db    3733360 TGGCGGAGCCGGTGCCGCCGGTGATGTGACCTTGGCCGTCAACCAGGGTGCCGGCGGTGA 3733301

Qy        818 TGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAG 877

Db    3733300 CGGCGGCAACGGCGGTGAAGTGGGCGTTGGCGGCAAGGGTGGGGCCGGCGGTGTTAGCGC 3733241

Qy        878 TGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGG 937

Db    3733240 GAACCCGGCCCTGAACGGTTCGGCCGGGGCGAACGGCACCGCGCCCACCAGCGGCGGCAA 3733181

Qy        938 CGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGG 997

Db    3733180 CGGTGGCAACGGAGGTGCCGGCGCCACCCCCACCGTCGCGGGAGAAAACGGCGGCGCCGG 3733121

Qy        998 GCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAA 1057

Db    3733120 TGGTAACGGCGGCCATGGCGGGTCGGTCGGTAACGGCGGTGCGGGTGGTGCCGGCGGAAA 3733061

Qy       1058 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 1117

Db    3733060 TGGCGTCGCCGGCACCGGCCTTGCCCTCAACGGCGGCAACGGCGGCAACGGCGGCATCGG 3733001

Qy       1118 GGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 1177

Db    3733000 CGGCAACGGCGGATCGGCGGCCGGCACGGGCGGGGACGGCGGCAAGGGCGGCAACGGGGG 3732941

Qy       1178 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 1237

Db    3732940 CGCCGGAGCCAACGGCCAAGACTTCTCCGCGTCCGCCAATGGCGCGAATGGCGGACAGGG 3732881

Qy       1238 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297

Db    3732880 CGGCAACGGCGGCAACGGCGGCATCGGCGGCAAGGGTGGTGACGCCTTCGCCACGTTCGC 3732821

Qy       1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357

Db    3732820 TAAGGCCGGCAACGGCGGTGCCGGCGGCAACGGCGGCAATGTGGGCGTTGCCGGCCAGGG 3732761

Qy       1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 1417

Db    3732760 TGGGGCCGGCGGCAAGGGCGCCATTCCAGCCATGAAGGGTGCGACCGGCGCCGATGGCAC 3732701

Qy       1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAAC 1448

Db    3732700 CGCACCCACCAGCGGCGGTGACGGCGGCAAC 3732670
```

RESULT 5
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 2.5%; Score 54.8; DB 3; Length 4411529;
Best Local Similarity 41.6%; Pred. No. 0.0093;
Matches 350; Conservative 0; Mismatches 492; Indels 0; Gaps 0;



```
Qy        607 GCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCT 666

Db    3741407 GCGACGGCGGCGTCGGCGGCAAGGGCGGTGCCGGCGGTAGCGGCGGCGCCGGCGGAAACG 3741348

Qy        667 GTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTA 726

Db    3741347 GATTCGACGCCGCCACCTTGGGTTCGCCCGGTGCCGATGGCGGTATGGGCGGCAACGGCG 3741288

Qy        727 ATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCA 786

Db    3741287 GCAAGGGCGGTGACGGCGGCAAGGCCGGTGATGGCGGAGCCGGTGCCGCCGGTGATGTGA 3741228

Qy        787 TTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAG 846

Db    3741227 CCTTGGCCGTCAACCAGGGTGCCGGCGGTGACGGCGGCAACGGCGGTGAAGTGGGCGTTG 3741168

Qy        847 GCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGA 906

Db    3741167 GCGGCAAGGGTGGGGCCGGCGGTGTTAGCGCGAACCCGGCCCTGAACGGTTCGGCCGGGG 3741108

Qy        907 ATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCA 966

Db    3741107 CGAACGGCACCGCGCCCACCAGCGGCGGCAACGGTGGCAACGGAGGTGCCGGCGCCACCC 3741048

Qy        967 GTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACC 1026

Db    3741047 CCACCGTCGCGGGAGAAAACGGCGGCGCCGGTGGTAACGGCGGCCATGGCGGGTCGGTCG 3740988

Qy       1027 GTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGC 1086

Db    3740987 GTAACGGCGGTGCGGGTGGTGCCGGCGGAAATGGCGTCGCCGGCACCGGCCTTGCCCTCA 3740928

Qy       1087 AAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGC 1146

Db    3740927 ACGGCGGCAACGGCGGCAACGGCGGCATCGGCGGCAACGGCGGATCGGCGGCCGGCACGG 3740868

Qy       1147 CCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCT 1206

Db    3740867 GCGGGGACGGCGGCAAGGGCGGCAACGGGGGCGCCGGAGCCAACGGCCAAGACTTCTCCG 3740808

Qy       1207 TCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTA 1266

Db    3740807 CGTCCGCCAATGGCGCGAATGGCGGACAGGGCGGCAACGGCGGCAACGGCGGCATCGGCG 3740748

Qy       1267 GAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGG 1326

Db    3740747 GCAAGGGTGGTGACGCCTTCGCCACGTTCGCTAAGGCCGGCAACGGCGGTGCCGGCGGCA 3740688

Qy       1327 GCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCT 1386

Db    3740687 ACGGCGGCAATGTGGGCGTTGCCGGCCAGGGTGGGGCCGGCGGCAAGGGCGCCATTCCAG 3740628

Qy       1387 TGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCA 1446

Db    3740627 CCATGAAGGGTGCGACCGGCGCCGATGGCACCGCACCCACCAGCGGCGGTGACGGCGGCA 3740568

Qy       1447 AC 1448

Db    3740567 AC 3740566
```

RESULT 6
US-09-252-991A-10929
; Sequence 10929, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10929
; LENGTH: 1221
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10929

Query Match 2.2%; Score 47.8; DB 4; Length 1221;
Best Local Similarity 50.5%; Pred. No. 0.0078;
Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

Qy 1097 CCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGT 1156
Db 345 CGACGGCAACATCCGCGTCGTGGAAATCAGCAGCGACGATGCCTGGGTCCGCGACACCGG 404

Qy 1157 CCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAA 1216
Db 405 CCCGACCTTCGTCATCGACGACAAGGGCGATGTACGCGGCGTCGACTGGGGCTTCAAC-- 462

Qy 1217 CATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGG 1276
Db 463 ----GCCTGGGGCGGCTTCGAAGGCGGCCTGTACTTCCCCTGGCAGCGCGACGACCAGGT 518

Qy 1277 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 1336
Db 519 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCCCGCTACCGCACCGACGACTTCGTCCT 578

Qy 1337 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATC 1385
Db 579 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 627

RESULT 7
US-09-252-991A-11061
; Sequence 11061, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11061
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11061

Query Match 2.2%; Score 47.8; DB 4; Length 1269;
Best Local Similarity 50.5%; Pred. No. 0.008;
Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

Qy 1097 CCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGT 1156
Db 316 CGACGGCAACATCCGCGTCGTGGAAATCAGCAGCGACGATGCCTGGGTCCGCGACACCGG 375

Qy 1157 CCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAA 1216
Db 376 CCCGACCTTCGTCATCGACGACAAGGGCGATGTACGCGGCGTCGACTGGGGCTTCAAC-- 433

Qy 1217 CATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGG 1276
Db 434 ----GCCTGGGGCGGCTTCGAAGGCGGCCTGTACTTCCCCTGGCAGCGCGACGACCAGGT 489

Qy 1277 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 1336
Db 490 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCCCGCTACCGCACCGACGACTTCGTCCT 549

Qy 1337 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATC 1385
Db 550 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 598

RESULT 8
US-09-252-991A-11134/c
; Sequence 11134, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11134
; LENGTH: 1641
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11134

Query Match 2.2%; Score 47.8; DB 4; Length 1641;
Best Local Similarity 50.5%; Pred. No. 0.0092;
Matches 146; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

Qy 1097 CCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGT 1156
Db 1399 CGACGGCAACATCCGCGTCGTGGAAATCAGCAGCGACGATGCCTGGGTCCGCGACACCGG 1340

Qy 1157 CCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAA 1216
Db 1339 CCCGACCTTCGTCATCGACGACAAGGGCGATGTACGCGGCGTCGACTGGGGCTTCAAC-- 1282

Qy 1217 CATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGG 1276
Db 1281 ----GCCTGGGGCGGCTTCGAAGGCGGCCTGTACTTCCCCTGGCAGCGCGACGACCAGGT 1226

Qy 1277 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 1336
Db 1225 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCCCGCTACCGCACCGACGACTTCGTCCT 1166

Qy 1337 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATC 1385
Db 1165 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTGATCACCACC 1117

RESULT 9
US-09-252-991A-10991
; Sequence 10991, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 10991
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10991

  Query Match             2.1%;  Score 46.8;  DB 4;  Length 501;
  Best Local Similarity   50.7%;  Pred. No. 0.0089;
  Matches  142;  Conservative    0;  Mismatches  132;  Indels    6;  Gaps    1;

Qy       1097 CCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGT 1156
             | | || ||  ||||| ||  ||| |||  |     | |   ||   | | |  | | |
Db        227 CGACGGCAACATCCGCGTCGTGGAAATCAGCAGCGACGATGCCTGGGTCCGCGACACCGG 286

Qy       1157 CCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAA 1216
              ||  ||  || ||| || ||||   | | |||  |  |    |  ||       |||||| 
Db        287 CCCGACCTTCGTCATCGACGACAAGGGCGATGTACGCGGCGTCGACTGGGGCTTCAAC-- 344

Qy       1217 CATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGG 1276
              ||||   || ||| |  || |  ||   ||||| |      ||    || || | 
Db        345 ----GCCTGGGGCGGCTTCGAAGGCGGCCTGTACTTCCCCTGGCAGCGCGACGACCAGGT 400

Qy       1277 AGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCAC 1336
              |  | |   |||   ||| || | |   ||| |  ||      | || | |   | |
Db        401 GGCACGCAAGATCCTCGAGATCGAACGGCGCGCCCGCTACCGCACCGACGACTTCGTCCT 460

Qy       1337 AGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTG 1376
              | ||| |    || | || ||||||| || || | ||||
Db        461 CGAGGGCGGCTCGATCCACGTCGACGGCGAAGGCACGCTG 500


RESULT 10
US-09-621-976-16656
; Sequence 16656, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 16656
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-16656

  Query Match             2.0%;  Score 43.8;  DB 4;  Length 430;
  Best Local Similarity   16.5%;  Pred. No. 0.051;
  Matches   51;  Conservative  123;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        915 AACTTCCTTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
              |:| :::::: :|  |:||    :  :: | :::   | ::  ::: ::|:   :    :
Db         59 AMCCYYYKKKGSCSRAMCCCCYTYCYKSCCSSYKGSYTTKRAMMWKRRSCYTYSRRRMYY 118

Qy        975 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 1034
              :::   : ::   :|  |  |: :: :::|: |:|| :::: |::: :  :  ::||:
Db        119 YWRSYWMMRSMWKGSCCCCSGGSCYKKKKKGKKGSCCMRSYWMCCYYYKRARRMMWKGGS 178

Qy       1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 1094
              :: |::::: :  | : :::::  :  :::|:: :|  :::: | |  :: |: ::: |
Db        179 CMMYTKRMMRRMCCCCMRRRSMRRRRMCMWKGSYTYCYKSSSMMCMARRWKRARGKKRMC 238

Qy       1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 1154
              :   |: :::  |:: :          || ||| |   | |  | | ||| | |
Db        239 CYTKGGGRMMRKYCCMRKKGRRACCTGTTCACCTGCCGTGTCTGCCAGAAGGCCTTCACC 298

Qy       1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
               ||       |  : |:|::| :::::  :   : :: ::::: :  ::: :  |:::  :
Db        299 TACCAGCGCATGYKRAMCSSCMMMWKRARKKKYMMMAMSRMSKYMARRRGSMMCYYYKSM 358

Qy       1215 AACATTGCC 1223
              :  : ::::
Db        359 MSKWMYKSS 367


RESULT 11
US-09-252-991A-547
; Sequence 547, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 547
;   LENGTH: 1176
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-547

  Query Match             2.0%;  Score 43.8;  DB 4;  Length 1176;
  Best Local Similarity   47.6%;  Pred. No. 0.089;
  Matches  166;  Conservative    0;  Mismatches  177;  Indels    6;  Gaps    1;

Qy       1112 CTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCAC 1171
              |||| |   ||| |||   |    | ||    |   || |       ||  | || ||||
Db        180 CTTCTGCTTCATGGCCGACCAGCCGATCCTGTTCTTCTTCAAGCCGGGCCAGTTCGTCAC 239

Qy       1172 CGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAG 1231
              |   ||  | || |  |||   |||| || |||   || | ||| ||| ||  |
Db        240 CCTGGAGCTGGAGATCGACGGCGAGCCGGTGATGCGCTCCTACACCATCTCCAGCTCGCC 299

Qy       1232 CTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGA 1291
              |||      || || || ||| |  |  ||||    | | ||  || | |   |  | |
Db        300 CTCGGTGCCCTACAGCTTCTCCATCACCATCAAGCGGGTGCCGGGCGGGCGGGTTTCCAA 359

Qy       1292 GGAGCT------CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGT 1345
              |||      ||| | |   || | |  ||||    ||  |  ||  | |  ||  |
Db        360 CTGGCTGCACGACAACCTCAAGGAAGGCCAGGAGCTACCGGTGCACGGTCCGGTCGGCCT 419

Qy       1346 GGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCAT 1405
              | | | || | |||||       |      ||  ||  || | |||    || |   |||
Db        420 GTTCAACGCCATCGACTTCCCGGCCGACAAGGTGCTGTTCCTCTCCGGCGGGGTCGGCAT 479

Qy       1406 GGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
              | | |  | ||||| |    ||  |       | |  ||||  ||||
Db        480 CACCCCGGTGATGTCCATGGCGCGCTGGTTCTTCGACACCAACGCCAAC 528


RESULT 12
US-09-252-991A-531
```

; Sequence 531, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 531
; LENGTH: 1983
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-531

Query Match 2.0%; Score 43.8; DB 4; Length 1983;
Best Local Similarity 47.6%; Pred. No. 0.12;
Matches 166; Conservative 0; Mismatches 177; Indels 6; Gaps 1;

```
Qy       1112 CTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCAC 1171
             |||| |   ||| |||   |   | ||   |  || |          ||  | || ||||
Db        205 CTTCTGCTTCATGGCCGACCAGCCGATCCTGTTCTTCTTCAAGCCGGGCCAGTTCGTCAC 264

Qy       1172 CGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAG 1231
             |   ||  | || |  |||   |||| || |||   || | ||| |||  ||  |
Db        265 CCTGGAGCTGGAGATCGACGGCGAGCCGGTGATGCGCTCCTACACCATCTCCAGCTCGCC 324

Qy       1232 CTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGA 1291
             |||     || || || ||| |  |  ||||    | | |||  || |  |   |  | |
Db        325 CTCGGTGCCCTACAGCTTCTCCATCACCATCAAGCGGGTGCCGGGCGGGCGGGTTTCCAA 384

Qy       1292 GGAGCT------CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGT 1345
              |||      ||| | |   || | |  |||||    || |  ||  | |  ||  |
Db        385 CTGGCTGCACGACAACCTCAAGGAAGGCCAGGAGCTACCGGTGCACGGTCCGGTCGGCCT 444

Qy       1346 GGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCAT 1405
             | | | || | |||||      |      ||  ||  || | |||   || |    |||
Db        445 GTTCAACGCCATCGACTTCCCGGCCGACAAGGTGCTGTTCCTCTCCGGCGGGGTCGGCAT 504

Qy       1406 GGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
              | | |  | |||||  |    ||  |       | |  ||||  ||||
Db        505 CACCCCGGTGATGTCCATGGCGCGCTGGTTCTTCGACACCAACGCCAAC 553
```

RESULT 13
US-09-252-991A-5938
; Sequence 5938, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5938
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5938

Query Match 2.0%; Score 43.4; DB 4; Length 1596;
Best Local Similarity 45.3%; Pred. No. 0.13;
Matches 158; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

```
Qy       1083 CTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCC 1142
             |||||    ||| | |  ||   |  ||    || || ||  ||   || |||| | ||
Db        1030 CTGCAGCACAGCGCGCCCGGAGTGCGCCTGCGCCTGGTCAATGCGGAACGCAAGCTGTCG 1089

Qy       1143 ATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAG 1202
              ||     ||| | | ||  |   |||      || | | | | | ||| |||| |  |||
Db        1090 GTGGAGGCCCTGGCCAGCGGGCGGATCGATTTCGCCCTCGGCTACGACGAGGAACACGAG 1149

Qy       1203 ATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATC 1262
              | |   | |  |||       ||    ||  |  ||| |||||  | |   ||
Db        1150 CGCCTGCCCGAAGGCATCCAGGCCCATGACTGGTTCGCCGACCGCTACGTGGTGGTAGCG 1209

Qy       1263 CGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCT 1322
              ||  | || |||  |  || |  |  | | |              ||| ||      || |
Db        1210 CGGCGCGACCACCCACGCCTCGCCGGGGCGCCGACGCTGGAGGGCTACCTGGCCGAACGG 1269

Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382
              | |   ||  ||  | |   | |  |  |  |       ||| || ||||||  || |
Db        1270 CATGCGGTGGTGACGCCGTGGAACGAGGACAGCGGGGTGATCGACCGGCTGCTGGCCCGC 1329

Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGG 1431
              |       | | | ||     ||  || | || | | | ||  |  ||
Db        1330 TCCGGCCTGCGTCGCGAAGTGGCGGTGCAACTGCCGACGGTGCTGGCGG 1378
```

RESULT 14
US-09-252-991A-5894/c
; Sequence 5894, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5894
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5894

Query Match 2.0%; Score 43.4; DB 4; Length 1674;
Best Local Similarity 45.3%; Pred. No. 0.14;
Matches 158; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

```
Qy       1083 CTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCC 1142
             |||||    ||| | |  ||   |  ||    || || ||  ||   || |||| | ||
Db         750 CTGCAGCACAGCGCGCCCGGAGTGCGCCTGCGCCTGGTCAATGCGGAACGCAAGCTGTCG 691

Qy       1143 ATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAG 1202
              ||     ||| | | ||  |   |||      || | | | | | ||| |||| |  |||
Db         690 GTGGAGGCCCTGGCCAGCGGGCGGATCGATTTCGCCCTCGGCTACGACGAGGAACACGAG 631

Qy       1203 ATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATC 1262
              | |   | |  |||       ||    ||  |  ||| |||||  | |   ||
Db         630 CGCCTGCCCGAAGGCATCCAGGCCCATGACTGGTTCGCCGACCGCTACGTGGTGGTAGCG 571

Qy       1263 CGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCT 1322
              ||  | || |||  |  || |  |  | | |              ||| ||      || |
Db         570 CGGCGCGACCACCCACGCCTCGCCGGGGCGCCGACGCTGGAGGGCTACCTGGCCGAACGG 511

Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382
              | |   ||  ||  | |   | |  |  |  |       ||| || ||||||  || |
Db         510 CATGCGGTGGTGACGCCGTGGAACGAGGACAGCGGGGTGATCGACCGGCTGCTGGCCCGC 451
```

```
Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGG 1431
Db        450 TCCGGCCTGCGTCGCGAAGTGGCGGTGCAACTGCCGACGGTGCTGGCGG 402
```

RESULT 15
US-09-489-039A-2746/c
; Sequence 2746, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2746
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2746

Query Match 2.0%; Score 43; DB 4; Length 486;
Best Local Similarity 45.4%; Pred. No. 0.09;
Matches 154; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

```
Qy        265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
Db        342 ACAGTCAGCCCACCCACGCGGTGGCGGTGTTCGATGATGAGGATCGCGCCCACGGCTGGC 283

Qy        325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
Db        282 GCCACCAGCACCTGCCGGAGTACAAAGCCGGCCGCGCGCCGATGCCGGAGACGCTGGTGG 223

Qy        385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
Db        222 CCGAGATGCCGGCGCTGCGGGCCGCCTTCGAACAGCGGGGGATCCGCTGCTGGGCCTCGC 163

Qy        445 GGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
Db        162 CGGGCAGCGAAGCAGACGATCTCGCCGCCACCCTGGCGGTGAAGGTGGCCCAGGCGGGCC 103

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
Db        102 ATCAGGCCACCATCGTCTCCACCGACAAAGGCTACTGCCAGCTGCTGTCGCCCACCATCC 43

Qy        565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCC 603
Db         42 GCATCCGCGACTACTTTCAGAAGCGCTGGCTGGACGCCC 4
```

Search completed: August 9, 2004, 21:19:09
Job time : 198 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 21:19:23 ; Search time 985 Seconds
(without alignments)
10836.686 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2043.8	93.9	2089	12	US-10-332-929A-17	Sequence 17, Appl
2	1893	87.0	2263	9	US-09-765-231A-30	Sequence 30, Appl
3	1808.8	83.1	2589	12	US-10-332-929A-1	Sequence 1, Appli
4	1807.8	83.0	2039	12	US-10-332-929A-16	Sequence 16, Appl
5	1807.8	83.0	2145	12	US-10-332-929A-18	Sequence 18, Appl
6	1794.4	82.4	2507	16	US-10-295-027-182	Sequence 182, App
7	1794.4	82.4	2507	17	US-10-188-832-57	Sequence 57, Appl
8	1687.8	77.5	1849	9	US-09-822-830A-88	Sequence 88, Appl
9	1408.8	64.7	2178	16	US-10-295-027-180	Sequence 180, App
10	1408.8	64.7	2178	17	US-10-188-832-55	Sequence 55, Appl
11	1345.8	61.8	1501	15	US-10-037-270-1032	Sequence 1032, Ap
12	1345.8	61.8	1501	16	US-10-117-722-1032	Sequence 1032, Ap
13	1025.6	47.1	1143	10	US-09-892-877-18	Sequence 18, Appl
14	1025.6	47.1	1143	10	US-09-948-783-18	Sequence 18, Appl
15	788	36.2	789	17	US-10-641-643-484	Sequence 484, App
16	687.2	31.6	4794	16	US-10-295-027-184	Sequence 184, App
17	687.2	31.6	4794	17	US-10-188-832-59	Sequence 59, Appl
18	429	19.7	465	10	US-09-918-995-19176	Sequence 19176, A
19	397.4	18.3	447	10	US-09-918-995-29322	Sequence 29322, A
c 20	229	10.5	492	13	US-10-240-425-1387	Sequence 1387, Ap
21	228.4	10.5	399	13	US-10-085-783A-21919	Sequence 21919, A
22	228.4	10.5	399	16	US-10-242-535A-21919	Sequence 21919, A
23	187	8.6	500	13	US-10-085-783A-20597	Sequence 20597, A
24	187	8.6	500	16	US-10-242-535A-20597	Sequence 20597, A
25	160	7.3	719	13	US-10-027-632-173199	Sequence 173199,
26	160	7.3	719	16	US-10-027-632-173199	Sequence 173199,
c 27	159.6	7.3	740	13	US-10-027-632-149888	Sequence 149888,
c 28	159.6	7.3	740	16	US-10-027-632-149888	Sequence 149888,
29	109	5.0	707	17	US-10-641-643-209	Sequence 209, App
30	79	3.6	572	13	US-10-027-632-244176	Sequence 244176,
31	79	3.6	572	16	US-10-027-632-244176	Sequence 244176,
32	77	3.5	325	13	US-10-085-783A-54629	Sequence 54629, A
33	77	3.5	325	16	US-10-242-535A-54629	Sequence 54629, A
34	60	2.8	60	10	US-09-908-975-16348	Sequence 16348, A
35	54	2.5	1443	15	US-10-156-761-4460	Sequence 4460, Ap
c 36	54	2.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
c 37	50.2	2.3	990	15	US-10-029-386-22613	Sequence 22613, A
38	49.8	2.3	1443	15	US-10-156-761-4458	Sequence 4458, Ap
39	49.8	2.3	1692	17	US-10-437-963-53149	Sequence 53149, A
40	49.6	2.3	1176	15	US-10-156-761-7518	Sequence 7518, Ap
41	49.4	2.3	1161	15	US-10-156-761-3651	Sequence 3651, Ap
c 42	49.2	2.3	520	15	US-10-184-644-332	Sequence 332, App
c 43	49.2	2.3	520	15	US-10-184-634-332	Sequence 332, App
44	49	2.3	785	13	US-10-425-114-35194	Sequence 35194, A
45	48.2	2.2	1815	9	US-09-758-269-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-332-929A-17
; Sequence 17, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 2089
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-332-929A-17

Query Match 93.9%; Score 2043.8; DB 12; Length 2089;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2045; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy      191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy      251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy      311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282
```

```
Qy       371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
Db       283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy       431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
Db       343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy       491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
Db       403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy       551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
Db       463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy       611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
Db       523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy       671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
Db       583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy       731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
Db       643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy       791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
Db       703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy       851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
Db       763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy       911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
Db       823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy       971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
Db       883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy      1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
Db       943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy      1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
Db      1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy      1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
Db      1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy      1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db      1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy      1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db      1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy      1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db      1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy      1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
Db      1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy      1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
Db      1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy      1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
Db      1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy      1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
Db      1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy      1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
Db      1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy      1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
Db      1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy      1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
Db      1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy      1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
Db      1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy      1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
Db      1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy      1931 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1990
Db      1843 AGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAG 1902

Qy      1991 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 2050
Db      1903 CCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTTGATC 1962

Qy      2051 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2110
Db      1963 AGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCTATTGTATTAAGCTATT 2022

Qy      2111 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2170
Db      2023 AATACATTAAGATTTGGGGGTGCTACCTTACATAATAAATTCCCATTTCCTCTTGAAAAA 2082

Qy      2171 AAAAAAA 2177
Db      2083 AAAAAAA 2089
```

RESULT 2
US-09-765-231A-30
; Sequence 30, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
; APPLICANT: Searle/Monsanto
; APPLICANT: Phippard, Deborah
; APPLICANT: Vasanthakamur, Geetha
; APPLICANT: Dotson, Stanton
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,
; TITLE OF INVENTION: vectors, and cells
; FILE REFERENCE: SO-3221 PR
; CURRENT APPLICATION NUMBER: US/09/765,231A
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 30
; LENGTH: 2263
; TYPE: DNA
; ORGANISM: Homo sapiens

```
;   FEATURE:
;   NAME/KEY: 1-2263
;   LOCATION: unknown
;   OTHER INFORMATION: unsure at all n locations
US-09-765-231A-30

  Query Match             87.0%;  Score 1893;  DB 9;  Length 2263;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1918;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 212

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 272

Qy        139 ACCCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 332

Qy        199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 392

Qy        259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 452

Qy        319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATG 512

Qy        379 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACG 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 ACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGTAACCCCTACTACG 572

Qy        439 CGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACG 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        573 CGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACG 632

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 692

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 618
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCA 752

Qy        619 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 678
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        753 ACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAA 812

Qy        679 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 738
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 AGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTG 872

Qy        739 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 798
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 ATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCGCTCA 932

Qy        799 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        933 GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGG 992

Qy        859 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 918
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        993 GCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT 1052

Qy        919 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 978
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1053 TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG 1112

Qy        979 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1113 ACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAG 1172

Qy       1039 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1173 TGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCC 1232

Qy       1099 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1233 ATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCC 1292

Qy       1159 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1293 GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACA 1352

Qy       1219 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1353 TTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAG 1412

Qy       1279 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1413 ACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAG 1472

Qy       1339 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 GGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAG 1532

Qy       1399 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 AGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGC 1592

Qy       1459 TGCGAGTGGTGCC-ACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1517
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1593 TGCGAGTGGTGCCAACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTC 1652

Qy       1518 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1653 TACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGT 1712

Qy       1578 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1637
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 GAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTG 1772

Qy       1638 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1697
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1773 ACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTG 1832

Qy       1698 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1757
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1833 CTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGT 1892

Qy       1758 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG-CATGGACACCAATGAATGCATCCA 1816
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1893 GGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCCATGGACACCAATGAATGCATCCA 1952

Qy       1817 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 1876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1953 GTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAG 2012

Qy       1877 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 1936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2013 GTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTG 2072

Qy       1937 CGTGG 1941
              |||||
Db       2073 CGTGG 2077


RESULT 3
US-10-332-929A-1
; Sequence 1, Application US/10332929A
```

```
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 2589
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2498)..(2498)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-332-929A-1

  Query Match             83.1%;  Score 1808.8;  DB 12;  Length 2589;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1810;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy       191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy       251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy       311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy       371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy       431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy       491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy       551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy       611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy       671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy       731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy       791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy       851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy       911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy       971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy      1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy      1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy      1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy      1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy      1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy      1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy      1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy      1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy      1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy      1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy      1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
            ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy      1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy      1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy      1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy      1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy      1931 AGCCTGCGTGGG 1942
            ||||||||||||
Db      1843 AGCCTGCGTGGG 1854
```

RESULT 4
US-10-332-929A-16
; Sequence 16, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
; APPLICANT: Akzo Nobel N.V.
; TITLE OF INVENTION: Novel extracellular matrix protein
; FILE REFERENCE: 2000.564
; CURRENT APPLICATION NUMBER: US/10/332,929A
; CURRENT FILING DATE: 2003-10-08
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 2039
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-332-929A-16

Query Match 83.0%; Score 1807.8; DB 12; Length 2039;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1809; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy   131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy   191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy   251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy   311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy   371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy   431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy   491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy   551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy   611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy   671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy   731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy   791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy   851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy   911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy   971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy  1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy  1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy  1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy  1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy  1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy  1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1243 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy  1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1362

Qy  1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy  1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy  1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy  1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy  1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy  1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy  1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy  1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy  1931 AGCCTGCGTGG 1941
             |||||||||||
```

```
Db       1843 AGCCTGCGTGG 1853


RESULT 5
US-10-332-929A-18
; Sequence 18, Application US/10332929A
; Publication No. US20040072286A1
; GENERAL INFORMATION:
;  APPLICANT: Akzo Nobel N.V.
;  TITLE OF INVENTION: Novel extracellular matrix protein
;  FILE REFERENCE: 2000.564
;  CURRENT APPLICATION NUMBER: US/10/332,929A
;  CURRENT FILING DATE:  2003-10-08
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 2145
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-332-929A-18

  Query Match             83.0%;  Score 1807.8;  DB 12;  Length 2145;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1809;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        131 GAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 190
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         43 GCGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCC 102

Qy        191 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 CATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCT 162

Qy        251 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGT 222

Qy        311 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCT 282

Qy        371 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACC 342

Qy        431 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGA 490
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        343 CTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGA 402

Qy        491 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 550
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 CGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGC 462

Qy        551 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 610
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 CACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGA 522

Qy        611 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 670
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGT 582

Qy        671 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 730
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGT 642

Qy        731 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 790
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 GGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCT 702

Qy        791 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 850
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 GGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGT 762

Qy        851 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 910
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGG 822

Qy        911 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 970
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 GCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGC 882

Qy        971 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 1030
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGA 942

Qy       1031 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1090
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 TGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAAT 1002

Qy       1091 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 GAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTC 1062

Qy       1151 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT 1122

Qy       1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1182

Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1242

Qy       1331 GGGCACAGGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1243 GGGCACAGGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1302

Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGGCAATCAGGGCTTCAACAA 1450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1303 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGGCAATCAGGGCTTCAACAA 1362

Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1510
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1363 CAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGT 1422

Qy       1511 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1570
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1423 CGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTA 1482

Qy       1571 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1630
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1483 CCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGT 1542

Qy       1631 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1690
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1543 GGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAA 1602

Qy       1691 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1750
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 CTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACT 1662

Qy       1751 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1810
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 GGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATG 1722

Qy       1811 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 CATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAG 1782

Qy       1871 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 CTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCAC 1842

Qy       1931 AGCCTGCGTGG 1941
```

```
                ||||||||||||
Db       1843 AGCCTGCGTGG 1853


RESULT 6
US-10-295-027-182
; Sequence 182, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 182
;   LENGTH: 2507
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-295-027-182

  Query Match             82.4%;  Score 1794.4;  DB 16;  Length 2507;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1795;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy       207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy       267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy       327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy       387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy       447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy       507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy       567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy       627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy       687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy       747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy       807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy       867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy       927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy       987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy      1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy      1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy      1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy      1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy      1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy      1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy      1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy      1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
```

```
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

RESULT 7
US-10-188-832-57
; Sequence 57, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002330US
; CURRENT APPLICATION NUMBER: US/10/188,832
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US 60/302,814
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/310,099
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/343,705
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 2507
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-188-832-57

Query Match 82.4%; Score 1794.4; DB 17; Length 2507;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 60

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 120

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 180

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 240

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 300

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 360

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 420

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 480

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 540

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 600

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 660

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 720

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 780

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 840

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 900

Qy       1047 GTCTATGGCAACCGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GTCTATGGCAACCGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 960

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1020

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1080

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1140

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1141 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1200

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1260

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1320

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1380

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1440

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1560

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1620

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1680

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1740

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1796
```

```
RESULT 8
US-09-822-830A-88
; Sequence 88, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 88
;   LENGTH: 1849
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1810,1813,1814
;   OTHER INFORMATION: n=a,c,g, or t
```

US-09-822-830A-88

```
  Query Match             77.5%;  Score 1687.8;  DB 9;  Length 1849;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1700;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 60

Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 120

Qy        359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 180

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        181 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGC 240

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 300

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 360

Qy        599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 420

Qy        659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 718
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 480

Qy        719 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 778
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 540

Qy        779 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 838
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 600

Qy        839 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 660

Qy        899 CAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 958
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATGAGAATGGGCCTAAATTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 720

Qy        959 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 780

Qy       1019 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 1078
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 840

Qy       1079 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 1138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 900

Qy       1139 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 960

Qy       1199 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1020
```

```
Qy       1259 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1080

Qy       1319 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1378
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1140

Qy       1379 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1200

Qy       1439 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1498
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1260

Qy       1499 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1558
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1320

Qy       1559 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1618
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1380

Qy       1619 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1440

Qy       1679 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1500

Qy       1739 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1798
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1560

Qy       1799 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1858
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1620

Qy       1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAA 1918
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1621 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTC-GGGCTACGAGCCCAA 1679

Qy       1919 CGAGGATGGCACAGCCTGCGTGG 1941
             |||||||||||||||||||||||
Db       1680 CGAGGATGGCACAGCCTGCGTGG 1702
```

RESULT 9
US-10-295-027-180
; Sequence 180, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 180
; LENGTH: 2178
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-180

Query Match 64.7%; Score 1408.8; DB 16; Length 2178;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 2; Indels 43; Gaps 1;

```
Qy        435 TACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGG 494
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 TACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGG 60

Qy        495 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 120

Qy        555 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 180

Qy        615 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 240

Qy        675 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 300

Qy        735 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCG 360

Qy        795 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 420

Qy        855 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 480

Qy        915 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 540

Qy        975 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 600

Qy       1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 660
```

```
Qy       1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 1154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 720

Qy       1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 780

Qy       1215 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 1274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 840

Qy       1275 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 1334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 900

Qy       1335 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 1394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 960

Qy       1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1020

Qy       1455 TGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1514
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 TGGCTGCGAGTGGTGCCACGCACCCGGGTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1080

Qy       1515 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1574
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1140

Qy       1575 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200

Qy       1635 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1694
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1260

Qy       1695 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAG 1754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGA- 1319

Qy       1755 TGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATC 1814
                                                 |||||||||||||||||
Db        1320 ----------------------------------------GACACCAATGAATGCATC 1337

Qy       1815 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1338 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1397

Qy       1875 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1398 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1457

Qy       1935 TGCGTGGG 1942
             ||||||||
Db        1458 TGCGTGGG 1465
```

```
RESULT 10
US-10-188-832-55
; Sequence 55, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
;  APPLICANT: Mack, David H.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions
;  TITLE OF INVENTION:  and Methods of Screening for Modulators of Bladder
;  TITLE OF INVENTION:  Cancer
;  FILE REFERENCE: 018501-002330US
;  CURRENT APPLICATION NUMBER: US/10/188,832
;  CURRENT FILING DATE:  2002-11-22
;  PRIOR APPLICATION NUMBER: US 60/302,814
;  PRIOR FILING DATE: 2001-07-03
;  PRIOR APPLICATION NUMBER: US 60/310,099
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/343,705
;  PRIOR FILING DATE: 2001-11-08
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/372,246
;  PRIOR FILING DATE: 2002-04-12
;  NUMBER OF SEQ ID NOS: 207
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
;   LENGTH: 2178
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-188-832-55

  Query Match             64.7%;  Score 1408.8;  DB 17;  Length 2178;
  Best Local Similarity   97.0%;  Pred. No. 0;
  Matches 1463;  Conservative    0;  Mismatches    2;  Indels   43;  Gaps    1;
```

```
Qy        435 TACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGG 494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGG 60

Qy        495 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACG 120

Qy        555 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAG 180

Qy        615 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGAC 240

Qy        675 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGC 300

Qy        735 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCG 794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGGCATTCTGGCG 360

Qy        795 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 420

Qy        855 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 480

Qy        915 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 540

Qy        975 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGC 600

Qy       1035 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGC 660

Qy       1095 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCT 720
```

```
Qy       1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 780

Qy       1215 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 1274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 840

Qy       1275 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 1334
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGC 900

Qy       1335 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 1394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCAT 960

Qy       1395 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1454
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAAC 1020

Qy       1455 TGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1514
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1021 TGGCTGCGAGTGGTGCCACGCACCCGGGTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTG 1080

Qy       1515 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1574
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTG 1140

Qy       1575 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1634
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAG 1200

Qy       1635 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1694
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCA 1260

Qy       1695 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAG 1754
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGA- 1319

Qy       1755 TGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATC 1814
                                                 ||||||||||||||||||
Db       1320 ----------------------------------------GACACCAATGAATGCATC 1337

Qy       1815 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 CAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTAC 1397

Qy       1875 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1934
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1398 AGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCC 1457

Qy       1935 TGCGTGGG 1942
             ||||||||
Db       1458 TGCGTGGG 1465
```

RESULT 11
US-10-037-270-1032
; Sequence 1032, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 1032
; LENGTH: 1501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(1468)
US-10-037-270-1032

Query Match 51.8%; Score 1345.8; DB 15; Length 1501;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        593 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGG 652
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db          1 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCTTGTGGTGTGGCCAGCCTCTTTGCCGG 60

Qy        653 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 120

Qy        713 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 180

Qy        773 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 240

Qy        833 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 300

Qy        893 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 360

Qy        953 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 420

Qy       1013 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 480

Qy       1073 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 540

Qy       1133 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 600

Qy       1193 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 660
```

```
Qy       1253 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 720

Qy       1313 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 1372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 780

Qy       1373 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 1432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 840

Qy       1433 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 1492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 900

Qy       1493 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 1552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 960

Qy       1553 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1020

Qy       1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1080

Qy       1673 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1140

Qy       1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1200

Qy       1793 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1260

Qy       1853 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1320

Qy       1913 GCCCAACGAGGATGGCACAGCCTGCGTGG 1941
             |||||||||||||||||||||||||||||
Db       1321 GCCCAACGAGGATGGCACAGCCTGCGTGG 1349
```

RESULT 12
US-10-117-722-1032
; Sequence 1032, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0

; SEQ ID NO 1032
; LENGTH: 1501
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(1468)
US-10-117-722-1032

Query Match 61.8%; Score 1345.8; DB 16; Length 1501;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy        593 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGG 652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACATCCTGAGCGATGAGGTCAACGTGGCTTGTGGTGTGGCCAGCCTCTTTGCCGG 60

Qy        653 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAA 120

Qy        713 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGA 180

Qy        773 CCTCTCCCGGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCTCCCGGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATA 240

Qy        833 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTT 300

Qy        893 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGT 360

Qy        953 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 1012
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCT 420

Qy       1013 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 1072
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCA 480

Qy       1073 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACC 540

Qy       1133 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCA 600

Qy       1193 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 1252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTT 660

Qy       1253 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 1312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGC 720

Qy       1313 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 1372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGAT 780

Qy       1373 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 1432
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGG 840

Qy       1433 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 1492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTT 900
```

```
Qy       1493 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 1552
Db        901 TGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCAT 960

Qy       1553 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1612
Db         961 CGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1020

Qy       1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1672
Db       1021 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1080

Qy       1673 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1732
Db       1081 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1140

Qy       1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1792
Db       1141 TCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTG 1200

Qy       1793 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1852
Db       1201 CATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATG 1260

Qy       1853 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1912
Db       1261 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1320

Qy       1913 GCCCAACGAGGATGGCACAGCCTGCGTGG 1941
Db       1321 GCCCAACGAGGATGGCACAGCCTGCGTGG 1349
```

RESULT 13
US-09-892-877-18
; Sequence 18, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: PZ028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1100)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-18

Query Match 47.1%; Score 1025.6; DB 10; Length 1143;
Best Local Similarity 98.5%; Pred. No. 1.4e-295;
Matches 1119; Conservative 0; Mismatches 9; Indels 8; Gaps 8;

```
Qy        815 GGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAG 874
Db          9 GGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA-CGTGGGCCCCATCCTCAGCAG 67

Qy        875 CAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCG 934
Db         68 CAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCG 127

Qy        935 GGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCA 994
Db        128 GGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCA 187

Qy        995 TGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGG 1054
Db        188 TGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGG 247

Qy       1055 CAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTT 1114
Db        248 CAACTGGAATGGCCCCCACCGCCTCTATCTGCAGATGAGCACCCATGGGAAGGTCCGCTT 307

Qy       1115 CC-GGGACATCGCC-TCACCCAAGTTCTCCATGCCCTCCCCTG-TCCGCACGGTCATCAC 1171
Db        308 CCGGGGACATCGCCTTCACCCAAGTTCTCCATGCCCTCCCCTGTTCCGCACGGTCATCAC 367

Qy       1172 C-GCCGACTTTGACAATGACCAGGAGCTGGAG-ATCTTCTTCAACAACATTGCCTACCGC 1229
Db        368 CGGCCGACTTTGACAATGACCAGGAGCTGGAGAATCTTCTTCAACAACATTGCCTACCGC 427

Qy       1230 AGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATC 1289
Db        428 AGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATC 487

Qy       1290 GAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTG 1349
Db        488 GAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTG 547

Qy       1350 ACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCT 1409
Db        548 ACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCT 607

Qy       1410 CAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTG 1469
Db        608 CAACCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTG 667

Qy       1470 CCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG 1529
Db        668 CCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG 727

Qy       1530 AGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC 1589
Db        728 AGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC 787

Qy       1590 GTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT 1649
Db        788 GTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT 847

Qy       1650 GGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTC 1709
Db        848 GGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTC 907

Qy       1710 TACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTC 1769
Db        908 TACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTC 967

Qy       1770 TCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTG 1829
Db        968 TCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTG 1027

Qy       1830 TGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAAC 1889
Db       1028 TGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAAC 1087

Qy       1890 AAGAAGTGCAGT-CGGGGCTACGAG-CCCAACGAGGATGGCACAGCCTGCGTGGGC 1943
Db       1088 AAGAAGTGCAGTNCGGGGCTACGAGTCCCAACGAGGATGGCACATACGGGCTTGTC 1143
```

RESULT 14
US-09-948-783-18
; Sequence 18, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: PZ028P2

```
;  CURRENT APPLICATION NUMBER: US/09/948,783
;  CURRENT FILING DATE:  2001-09-10
;  PRIOR APPLICATION NUMBER: 60/231,846
;  PRIOR FILING DATE: 2000-09-11
;  PRIOR APPLICATION NUMBER: 09/892,877
;  PRIOR FILING DATE: 2001-06-28
;  PRIOR APPLICATION NUMBER: 09/437,658
;  PRIOR FILING DATE: 1999-11-10
;  PRIOR APPLICATION NUMBER: PCT/US99/09847
;  PRIOR FILING DATE: 1999-05-06
;  PRIOR APPLICATION NUMBER: 60/085,093
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,094
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,105
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,180
;  PRIOR FILING DATE: 1998-05-12
;  PRIOR APPLICATION NUMBER: 60/085,927
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,906
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,924
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,922
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,921
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,923
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,925
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,928
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,920
;  PRIOR FILING DATE: 1998-05-18
;  NUMBER OF SEQ ID NOS: 465
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1143
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1100)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-18

  Query Match             47.1%;  Score 1025.6;  DB 10;  Length 1143;
  Best Local Similarity   98.5%;  Pred. No. 1.4e-295;
  Matches 1119;  Conservative    0;  Mismatches    9;  Indels    8;  Gaps    8;

Qy        815 GGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAG 874
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          9 GGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA-CGTGGGCCCCATCCTCAGCAG 67

Qy        875 CAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCG 934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 CAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCG 127

Qy        935 GGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCA 994
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 GGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCA 187

Qy        995 TGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGG 1054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 TGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGG 247

Qy       1055 CAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTT 1114
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        248 CAACTGGAATGGCCCCCACCGCCTCTATCTGCAGATGAGCACCCATGGGAAGGTCCGCTT 307

Qy       1115 CC-GGGACATCGCC-TCACCCAAGTTCTCCATGCCCTCCCCTG-TCCGCACGGTCATCAC 1171
                ||| |||||||||| |||||||||||||||||||||||||| |||||||||||||||
Db        308 CCGGGGACATCGCCTTCACCCAAGTTCTCCATGCCCTCCCCTGTTCCGCACGGTCATCAC 367

Qy       1172 C-GCCGACTTTGACAATGACCAGGAGCTGGAG-ATCTTCTTCAACAACATTGCCTACCGC 1229
              | ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        368 CGGCCGACTTTGACAATGACCAGGAGCTGGAGAATCTTCTTCAACAACATTGCCTACCGC 427

Qy       1230 AGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATC 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 AGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATC 487

Qy       1290 GAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTG 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 GAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTG 547

Qy       1350 ACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCT 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 ACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCT 607

Qy       1410 CAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTG 1469
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 CAACCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTG 667

Qy       1470 CCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 CCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAG 727

Qy       1530 AGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 AGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCC 787

Qy       1590 GTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT 1649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 GTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGAT 847

Qy       1650 GGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTC 1709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 GGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTC 907

Qy       1710 TACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTC 1769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 TACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTC 967

Qy       1770 TCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTG 1829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 TCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTG 1027

Qy       1830 TGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAAC 1889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 TGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAAC 1087

Qy       1890 AAGAAGTGCAGT-CGGGGCTACGAG-CCCAACGAGGATGGCACAGCCTGCGTGGGC 1943
              |||||||||||| |||||||||||| |||||||||||||||||||  | |  | | |
Db       1088 AAGAAGTGCAGTNCGGGGCTACGAGTCCCAACGAGGATGGCACATACGGGCTTGTC 1143


RESULT 15
US-10-641-643-484
; Sequence 484, Application US/10641643
; Publication No. US20040077003A1
;   GENERAL INFORMATION:
;        APPLICANT: Cocks, Benjamin G.
;                   Susan G. Stuart
;                   Jeffrey J. Seilhamer
;        TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
;                            GENE EXPRESSION
;        NUMBER OF SEQUENCES: 1508
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;             STREET: 3174 PORTER DRIVE
;             CITY: PALO ALTO
;             STATE: CALIFORNIA
;             COUNTRY: USA
```

```
;           ZIP: 94304
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/10/641,643
;           FILING DATE: 14-Aug-2003
;           CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: <Unknown>
;           FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Zeller, Karen J.
;           REGISTRATION NUMBER: 37,071
;           REFERENCE/DOCKET NUMBER: PA-0001 US
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (650) 855-0555
;           TELEFAX: (650) 845-4166
;   INFORMATION FOR SEQ ID NO: 484:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 789 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       IMMEDIATE SOURCE:
;           LIBRARY: SYNORAB01
;           CLONE: 192279
;       SEQUENCE DESCRIPTION: SEQ ID NO: 484 :
US-10-641-643-484

  Query Match             36.2%;  Score 788;  DB 17;  Length 789;
  Best Local Similarity   99.9%;  Pred. No. 1.3e-224;
  Matches  788;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1058 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 1117
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCG 60

Qy       1118 GGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 1177
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGACATCGNCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGA 120

Qy       1178 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 1237
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTC 180

Qy       1238 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 1297
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCT 240

Qy       1298 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 1357
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTT 300

Qy       1358 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCT 360

Qy       1418 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCAC 420

Qy       1478 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGC 480

Qy       1538 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACA 540

Qy       1598 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 1657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGAT 600

Qy       1658 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCG 660

Qy       1718 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 1777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCA 720

Qy       1778 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 1837
             ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCG 780

Qy       1838 AGACAAGCC 1846
             |||||||||
Db        781 AGACAAGCC 789


Search completed: August  9, 2004, 23:33:53
Job time : 995 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:38:45 ; Search time 37.5 Seconds
(without alignments)
11168.477 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161652_29697/app_query.fasta_1.2375
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914958_@CGN_1_1_57_@runat_03082004_161652_29697 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	274.5	6.8	3570	2	T45025	mucin MUC5B, trach
	2	257	6.3	1367	1	S48478	glucan 1,4-alpha-g
	3	249	6.1	528	2	I47141	gastric mucin (clo
	4	242	6.0	2232	2	T34434	hypothetical prote
	5	231	5.7	825	2	T29634	hypothetical prote
	6	229	5.6	3020	2	A43932	mucin 2 precursor,
c	7	223.5	5.4	660	1	QQBE3	BHLF1 protein - hu
	8	214.5	5.3	1032	2	T34433	hypothetical prote
	9	213.5	5.3	797	1	VGBEX1	glycoprotein X pre
	10	209.5	5.2	867	2	T45463	membrane glycoprot
	11	206.5	5.1	866	2	T45462	membrane glycoprot
	12	205.5	5.1	1952	2	T48814	hypothetical prote
	13	203.5	5.0	543	2	S35047	mucin JUL7 - human
	14	203.5	5.0	796	2	T21460	hypothetical prote
c	15	199.5	4.8	1106	2	JQ0405	hypothetical 119.5
	16	198	4.9	1367	2	T33819	hypothetical prote
	17	197.5	4.9	907	1	QQBE21	membrane antigen g
	18	195.5	4.8	786	2	T16509	hypothetical prote
c	19	195	4.7	1246	2	JQ0406	hypothetical prote
	20	193	4.8	580	2	T43481	probable mucin DKF
	21	192	4.7	662	2	A45155	mucin FIM-C.1 - Af
	22	192	4.7	1106	2	JQ0405	hypothetical 119.5
	23	191.5	4.7	1151	2	T18535	high molecular mas
c	24	191.5	4.6	1184	2	S50832	atrophin-1 - human
	25	191.5	4.7	1199	2	A40670	nuclear envelope p
	26	191.5	4.7	1820	2	A55494	latent transformin
c	27	188	4.5	1184	2	G01763	atrophin-1 - human
	28	187.5	4.6	805	2	T25795	hypothetical prote
	29	186.5	4.6	507	2	T44768	antifreeze glycope
	30	186.5	4.6	1589	2	T42233	submaxillary mucin
	31	184.5	4.5	1791	2	T02345	hypothetical prote
	32	184	4.5	648	2	PC4395	mucin 3 - human (f
c	33	183	4.4	1952	2	T48814	hypothetical prote
	34	182.5	4.5	1414	1	S23809	collagen alpha 2(I
	35	182	4.5	1184	2	S50832	atrophin-1 - human
c	36	182	4.4	1344	1	A35175	mucin 1 precursor,
	37	181	4.5	503	2	S31940	starch-degrading e
	38	181	4.5	1184	2	G01763	atrophin-1 - human
	39	181	4.5	1777	2	T34369	hypothetical prote
	40	180	4.4	503	2	C97262	N-terminal domain
	41	179.5	4.4	898	2	T14764	hypothetical prote
	42	179.5	4.4	1042	1	CGCH1S	collagen alpha 1(I
	43	179.5	4.4	1299	2	T47182	hypothetical prote
	44	179.5	4.4	1464	1	CGHU1S	collagen alpha 1(I
	45	179	4.4	1161	2	S57180	probable membrane

ALIGNMENTS

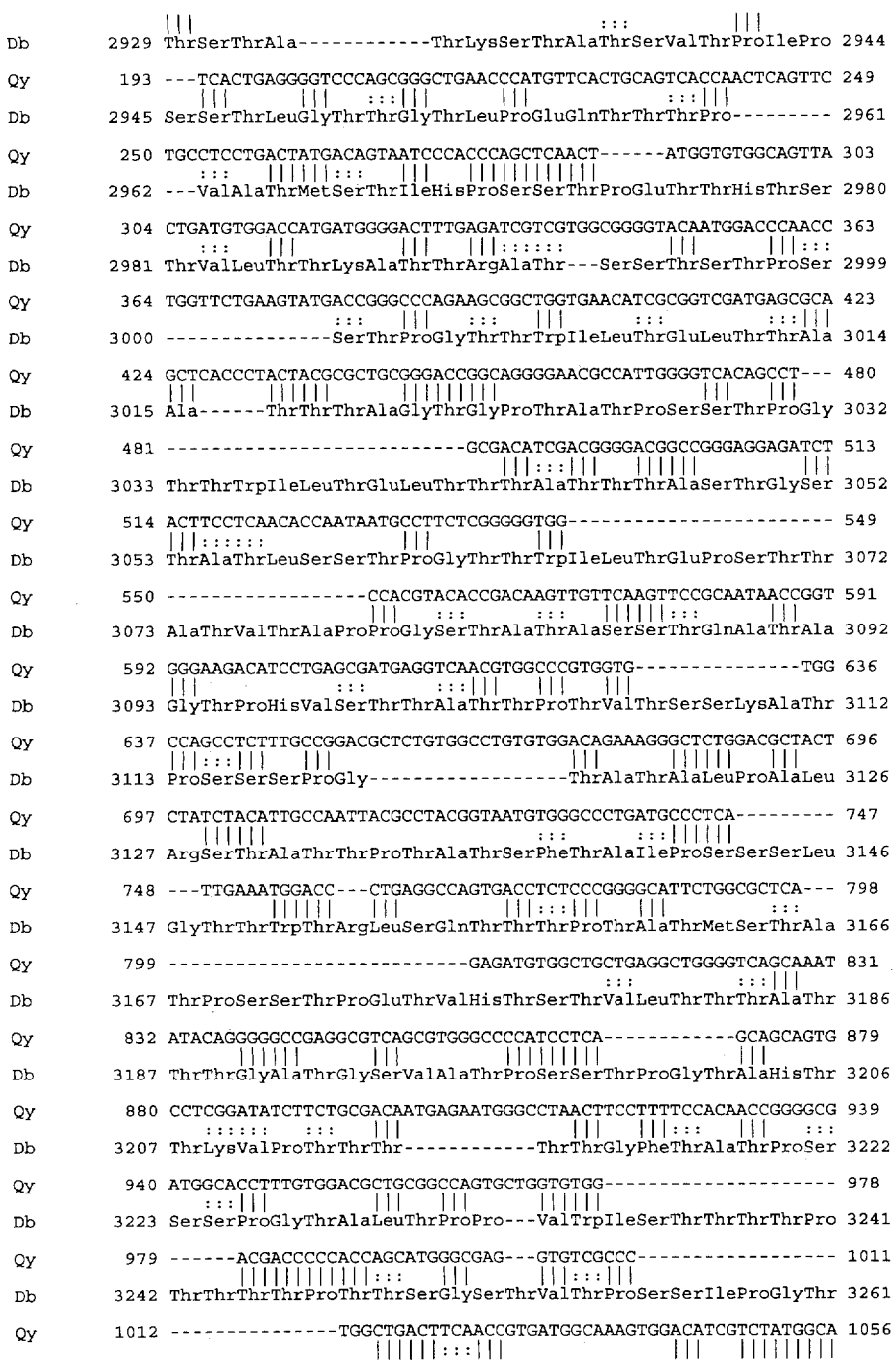
RESULT 1
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternate
A;Reference number: Z22899; MUID:97166151; PMID:9013550
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B

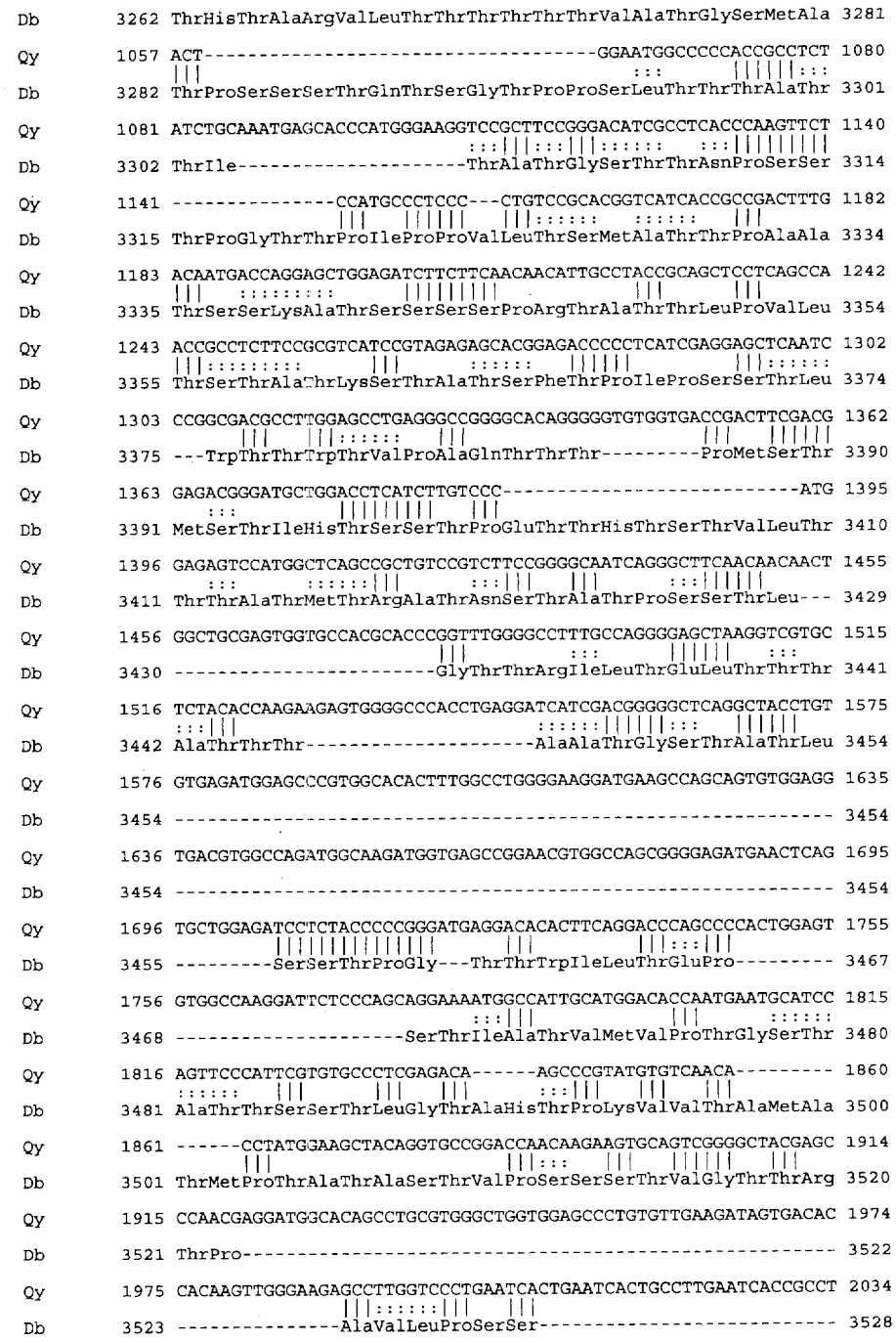
Alignment Scores:

Pred. No.:	4.26e-08	Length:	3570
Score:	274.50	Matches:	182
Percent Similarity:	34.16%	Conservative:	92
Best Local Similarity:	22.69%	Mismatches:	271
Query Match:	6.77%	Indels:	257
DB:	2	Gaps:	37

US-09-914-958B-35 (1-2177) x T45025 (1-3570)

```
Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
              |||||||||           |||         :::||||||           |||:::
Db       2890 ProGlyThrAlaProProProLysValLeuThrSerProAlaThrThrProThrAlaThr 2909

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
              :::                 ::::::   ||||||          |||      |||
Db       2910 SerSerLysAlaThrSer---SerSerSerProArgThrAlaThrThrLeuProValLeu 2928

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA------ 192
```





```
Qy       2035 GGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCT 2094
                                  ||||||      |||:::|||   :::   :::|||
Db       3529 ----------------------LeuProThrPheSerValSerThrValSerSerSer 3540

Qy       2095 ATTGTA 2100
                 ::::::
Db       3541 ValLeu 3542
```

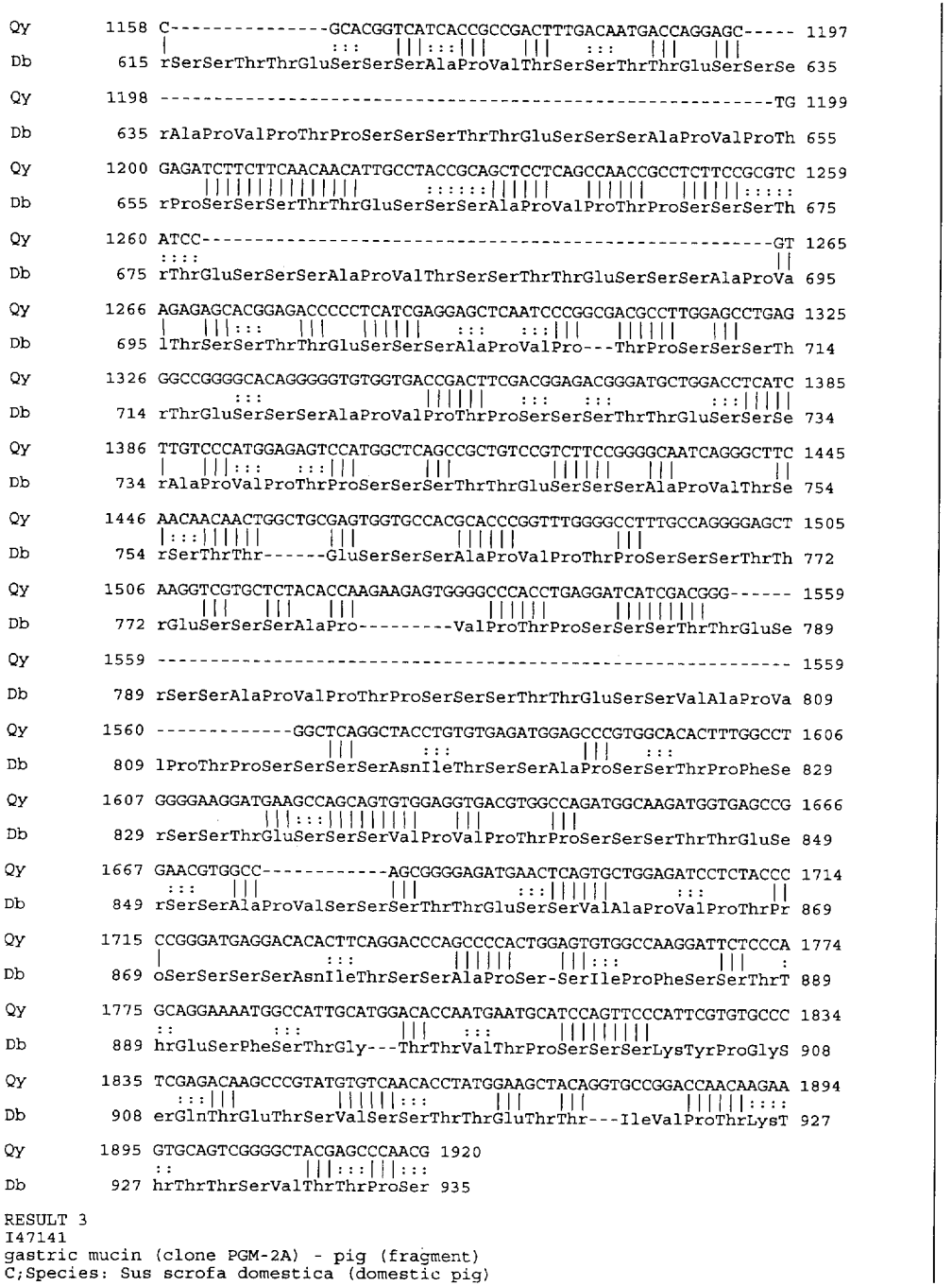
RESULT 2
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237; PMID:8710886
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Alignment Scores:

Pred. No.:	3.57e-07	Length:	1367
Score:	257.00	Matches:	182
Percent Similarity:	35.77%	Conservative:	72
Best Local Similarity:	25.63%	Mismatches:	274
Query Match:	6.34%	Indels:	183
DB:	1	Gaps:	30

US-09-914-958B-35 (1-2177) x S48478 (1-1367)

```
Qy         81 CCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTGAC 140
                ||:::|||||||||               ||||||        ||||||

Db        313 ProValProThrProSerSerSerThrThrGluSerSer---SerAlaProValProThr 331

Qy        141 CCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAG 200
                |||      |||                                        ||||||
Db        332 ProSerSerSer-------------------------------------ThrThrGlu 338

Qy        201 GGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGAC 260
                 |||      :::   |||:::   ::::::   |||   |||         |||
Db        339 SerSer------SerAlaProValThrSerSerThrThrGluSerSerSerAlaPro-Va 356

Qy        261 TATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGAT 320
                  |||      |||  |||||||||
Db        356 lProThr------ProSerSerSerThr------------------------------ 363

Qy        321 GGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGAC 380
                 |||      ||||||         |||         |||         |||   ::
Db        364 ----ThrGluSerSerSerAlaProValThrSerSerThrThrGlu------SerSerSe 380

Qy        381 CGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCG 440
                :  |||   :::       |||:::  |||  ||||||      ||||||
Db        380 rAlaProValThrSerSer---ThrThrGluSerSerSerAlaProValProThrProSe 399

Qy        441 CTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGC 500
                    |||         |||         |||  |||  |||||:::   |||
Db        399 rSerSerThr---------ThrGluSerSerSerAlaProValThrSerSerThrThr-- 415

Qy        501 CGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACC 560
                       :::||||||   |||:::      |||
Db        416 ----------GluSerSerSerAlaProValThrSerSer-----------------Th 426

Qy        561 GACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAAC 620
                 |||      ||||||:::   :::|||        |||                |||::
Db        426 rThrGluSerSerSerAlaProValThrSerSerThrThrGlu---------SerSerSe 443

Qy        621 GTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAG 680
                :  ||||||        ::::::        :::      ||||||   |||
Db        443 rAlaProValThr---SerSerThrThrGluSerSerSerAlaProValProThrProSe 462

Qy        681 GGCTCTGGACGCTACTCTATCTACATTGCCAATTACG---------------------- 717
                 :::                 |||:::   |||:::|||
Db        462 rSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSe 482

Qy        718 ----------CCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACC-------- 759
                         ||||||                 ||||||      |||
Db        482 rAlaProValProThr-----------------ProSerSerSerThrThrGluSerSe 496

Qy        760 -CTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCT 818
                 ||||||||||||
Db        496 rSerAlaProValThrSerSer-------------------------ThrThrGluSe 507

Qy        819 GGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGT 878
                  ||||||   :::           :::|||           ||||||||
Db        507 rSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProAl 527

Qy        879 GCCTCGGATATCTTCTGCGACAATGAGAATGGGC---------CTAACTTCCTTTTCCAC 929
                 |||      ||||||:::|||                  :::||||||   :::||
Db        527 aProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrTh 547

Qy        930 AACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCAC 989
                |      :::   ||||||::                                ||||||::
Db        547 rGluSerSerSerAlaProVal-------------------------ProThrProSe 558

Qy        990 CAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTC 1049
                :|||           |||         :::||||||   :::         |||:::
Db        558 rSerSerThrThrGluSer---------SerSerThrProValThrSerSerThrThrGl 575

Qy       1050 TATGGCAACTGGAATGGCCCCCACCGCCTCTATC------------TGCAAATGAGCACC 1097
                  ::::::        ||||||   |||                        |||||
Db        575 uSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaPr 595

Qy       1098 CATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTC 1157
                |:::         :::|||   |||:::         |||:::|||   |||     ||
Db        595 oValProThrProSerSerSerThrThrGluSerSerAlaProAlaProThrProSe 615
```

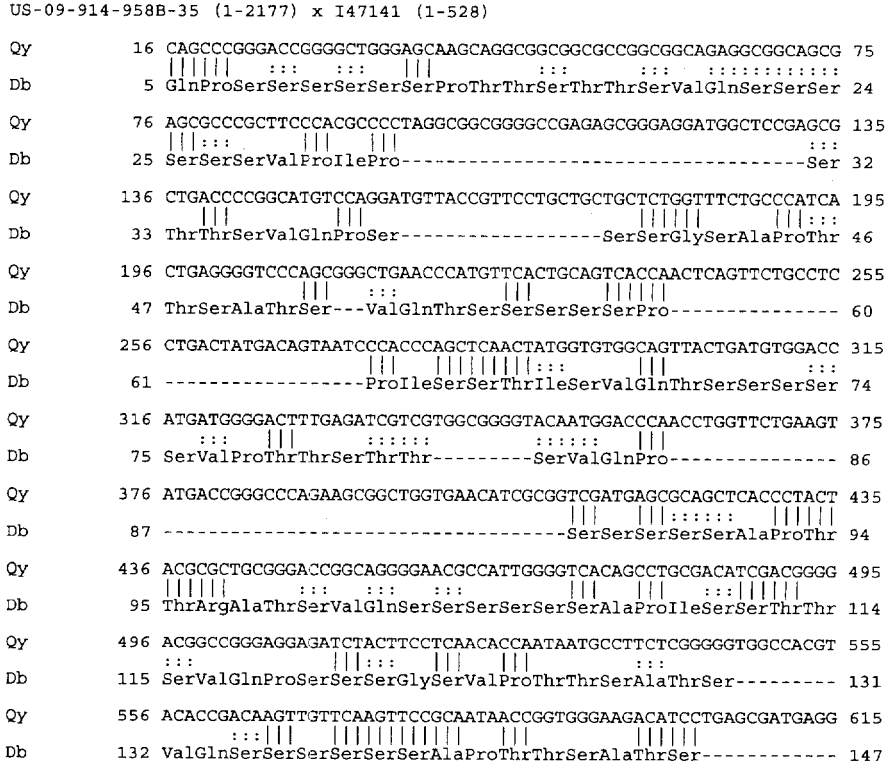


RESULT 3
I47141
gastric mucin (clone PGM-2A) - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C;Accession: I47141; S55315
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A;Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel
A;Reference number: I47141; MUID:94102478; PMID:7506218
A;Accession: I47141
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-528 <TUR>
A;Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A;Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A;Reference number: S55315; MUID:95275264; PMID:7755593
A;Accession: S55315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-528 <TU2>
A;Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
C;Superfamily: pig submaxillary mucin

Alignment Scores:

Pred. No.:	8.48e-07	Length:	528
Score:	249.00	Matches:	156
Percent Similarity:	36.90%	Conservative:	89
Best Local Similarity:	23.49%	Mismatches:	241
Query Match:	6.14%	Indels:	178
DB:	2	Gaps:	23

US-09-914-958B-35 (1-2177) x I47141 (1-528)



```
Db        148 --------------ValGlnProSerSer------------------------------ 152

Qy        676 GAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCC 735
                                 |||:::   ||||||:::   ||||||      :::
Db        153 --------------------SerSerSerProProIleSerSerThrVal------Ser 163

Qy        736 CTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGC 795
              :::   ||||||
Db        164 ValGlnProSer------------------------------------------------ 167

Qy        796 TCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCG 855
                                       |||::::::                 |||::::::
Db        168 ------------------------SerSerSerSerAlaProThrThrSerAlaThrSer 179

Qy        856 TGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCT--------------GCGACA 900
                |||||||||:::::::   |||   |||||||||                   :::
Db        180 ValGlnProSerSerSerSerSerProProIleSerSerThrValSerValGlnThrSer 199

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
                      :::   :::   |||||||||                   :::
Db        200 SerSerSerSerValProThrThrSerThrThrSerValGlnProSerSerSerSerSer 219

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
                 |||           |||:::      |||                |||    :::
Db        220 ValProThrThrSerAlaThrSerValArgSerSerSerSerSerSerThrProIlePro 239

Qy        1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
               ||||||         :::   :::||||||   :::         |||||||||   |||
Db         240 SerThrThrSerValGlnProSerSerSerSerSer---------AlaProThrThrSer 256

Qy        1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
                        |||          |||:::|||   ::::::|||   |||||||:::
Db         257 AlaThrSerValGlnPro---------SerSerSerSerSerThrProIleProSerThr 273

Qy        1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
                       |||:::   ||||||   ||||||   :::   |||
Db         274 ThrSerValGlnProSerSerSerSerSerAlaProThrThrSerAlaThrSerValGln 293

Qy        1201 AGATCTTCTTCAACAACATTGCCT---------ACCGCAGCTCCTCAGCCAACCGCCTCT 1251
                  |||||||||:::::::   |||          |||   :::   ||||||:::::::|||
Db         294 ProSerSerSerSerSerProProIleSerSerThrIleSerValGlnProSerSerSer 313

Qy        1252 TCCGCGTCATCCGTAGAGAGCACGGAGACC------CCCTCATCGAGGAGCTCAATCCCG 1305
               |||:::|||          ||||||   :::         |||||||||      |||   |||
Db         314 SerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySerAlaPro 333

Qy        1306 GCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAG 1365
                  |||                                          :::::::|||
Db         334 ThrThr-------------------------------------------SerAlaThrSer 339

Qy        1366 ACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCT 1425
                       :::||||||   |||       |||       :::|||       ||||||
Db         340 ValGlnProSerSerSerSerSerProProIleSerSerThrIleSerValGlnProSer 359

Qy        1426 TCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTG 1485
               |||   :::          :::|||:::|||                  :::   |||
Db         360 SerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 379

Qy        1486 GGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGA 1545
                  |||                                  ::::::      |||   |||:::
Db         380 AlaPro---------------------ThrThrSerAlaThrSerValGlnProSerSer 392

Qy        1546 GGATCATCGACG---GGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTG 1602
                |||||||                 ||||||   ||||||   |||
Db         393 SerSerSerValProThrThrSerAlaThrSerValArgSerSerSer------------ 408

Qy        1603 GCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGA 1662

Db         408 ------------------------------------------------------------ 408

Qy        1663 GCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATG 1722
               :::   :::   |||             |||             ||||||:::      :::

Db         409 SerSerSerThrProIleProThrThrThrSerValGlnProSerSerSerSerSerVal 428

Qy        1723 AGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAA 1782
                  |||         |||             ||||||         |||   |||
Db         429 ProThrThrSerAlaThr------------SerValGlnThrSerSerSerSerSerThr 444

Qy        1783 ATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGT------------ 1830
                        :::   |||   :::       |||||||||   |||
Db         445 ProIleProSerThrThrSerValGlnProSerSerSerSerSerAlaProThrThrSer 464

Qy        1831 ------GCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACA---GGTGCC 1881
                     :::::::::   |||         |||   |||:::   :::|||
Db         465 AlaThrSerValGlnProSerSerSerSerSerProProIleSerSerThrIleSerVal 484

Qy        1882 GGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
                  |||:::   |||:::         |||:::   |||         ||||||:::
Db         485 GlnProSerSerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSer 504

Qy        1942 GCTGGTGGAGCC 1953
                :::|||   |||
Db         505 SerGlySerAla 508

RESULT 4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Alignment Scores:
Pred. No.:              2.9e-06     Length:          2232
Score:                  242.00      Matches:         189
Percent Similarity:     31.48%      Conservative:    88
Best Local Similarity:  21.48%      Mismatches:      302
Query Match:            5.97%       Indels:          301
DB:                     2           Gaps:            30

US-09-914-958B-35 (1-2177) x T34434 (1-2232)

Qy           3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
                 |||   |||   |||:::                      :::   |||   |||
Db         458 GlySerSerThrProSerAlaSer---------SerSerSerAlaGlyThrAlaSerThr 474

Qy          63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGC---GGG 119
                  ||||||                  |||   :::   |||   :::   |||   |||
Db         475 IleSerGlySerThrGlySerThrAlaThrIleValProGlySerSerSerValGly 494

Qy         120 AGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTC 179
                        ||||||  :::|||:::   ||||||
Db         495 SerSerThrGlnSerAla-SerProSerSerProGlyThr---------------------- 507

Qy         180 TGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACC 239
                           ::::::   ||||||:::|||
Db         508 -------MetSerThrValSerGlyProThrGly-------------------------- 516

Qy         240 AACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATG-------- 291
                                 |||:::||||||:::|||   |||||||||
Db         517 ----------------SerThrValThrValValProGlySerSerThrSerProAlaPr 531
```

```
Db       1219 nPro------SerThrValArgGlySerThrSerSerGlySerThrValThrThrGlySe 1237

Qy       1803 AATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACC 1862

Db       1237 rThrGluGlySerSerThrSerGlySerSerSerAlaThrSerLeuSerSerSerSerPr 1257

Qy       1863 TATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Db       1257 oValProSerThrSerGlnSerProAsnProSerThrSerGlySerSerThrProThr 1276

RESULT 5
T29634
hypothetical protein C12D12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T29634
R;Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of C. elegans cosmid C12D12.
A;Reference number: Z20656
A;Accession: T29634
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-825 <NHA>
A;Cross-references: EMBL:U51998; PIDN:AAA96080.1; GSPDB:GN00028; CESP:C12D12.1
A;Experimental source: strain Bristol N2; clone C12D12
C;Genetics:
A;Gene: CESP:C12D12.1
A;Map position: X
A;Introns: 48/1; 86/3; 137/1; 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/1
C;Superfamily: Epstein-Barr virus membrane antigen gp350

Alignment Scores:
Pred. No.:              1.02e-05        Length:         825
Score:                  231.00          Matches:        119
Percent Similarity:     37.45%          Conservative:   54
Best Local Similarity:  25.76%          Mismatches:     154
Query Match:            5.70%           Indels:         135
DB:                     2               Gaps:           22

US-09-914-958B-35 (1-2177) x T29634 (1-825)

Qy         57 GGCGGCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGC 116

Db        447 GlyTyrGluGlyGlyAlaAsnLysProThrValVal---------------------- 458

Qy        117 GGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTG 176

Db        459 ------LeuAlaProProThrThrProGlyProThr---------------------- 468

Qy        177 CTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTC 236

Db        469 ------PheTyrProValThrThrValSerThrMetSerProProThrThrValThrVal 486

Qy        237 ACCAACTCAGTTCTGCCT---CCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGT 293

Db        487 ProThrThrProThrProValPro-ThrThrThrAsnThrProProAlaAsnProThrTh 506

Qy        294 GTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAAT 353

Db        506 r---------AlaThrProThrThrValGlyThr------SerLysGlnThrAsnThrIl 521

Qy        354 GGACCCAACCTGGTTCTGAAGTATGACCGGG----------CCCAGAAGCGG 395

Db        521 eSerPro---HisLeuSerThrIleThrGlySerIleValThrSerThrPro-------- 537

Qy        396 CTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAG 455

Db        538 -------ThrMetAlaProGlnThrSerAlaSerProThrThr---------------- 549

Qy        456 GGGAACGCCA---------TTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506

Db        550 ----ThrProThrHisThrThrAlaSerGlnProThrThrThrLysProValValThrTh 568

Qy        507 GAGATCT----------ACTTCCTCAACACCAATAATGCCTTCTCGGGGGTG 548

Db        568 rAsnSerValThrProSerThrGlyThrThrThrValProValProThrThrThrGlySe 588

Qy        549 GCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGC 608

Db        588 rProThrThrGlnThrThrAlaProValThrLysProThr------ValProSerSerTh 606

Qy        609 GATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGT 668

Db        606 rThrGlnThrAlaProProValThrThrProThrSerGlnPro------------ProVa 622

Qy        669 GTGGACAGAAAGGGCTCTGGACGCT---ACTCTATCTACATTGCCAATTACGCCTACGGT 725

Db        622 lThrThrThrSerLeuLeuThrThrLeuThrThrProThrValProValThrThrThrVa 642

Qy        726 AATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGC 785

Db        642 lVal---------------------------------------------------------- 643

Qy        786 ATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGA 845

Db        643 ---------------------------------------------------------- 643

Qy        846 GGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAG 905

Db        644 ----------------ProSerSerAlaThrValProThrThrProProThrThr----- 656

Qy        906 AATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCC 965

Db        657 -------ValThrValAlaAlaThrThrThrSerLysAlaProValValThrThrSerPr 674

Qy        966 AGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAAC 1025

Db        674 oThrLeuAlaProThrSerProThrLysLeuProThrSerProPro---------SerTh 691

Qy       1026 CGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCC-------- 1077

Db        691 rVal-----------------------------GlyThrSerProThrAlaProAlaAs 701

Qy       1078 ----TCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCT---CA 1130

Db        701 nLeuThrThrProThrThrAlaProValAsnProThrSerSerThrThrAlaProThrAl 721

Qy       1131 CCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGAC 1190

Db        721 aProValAsnProThrSerProThrThrAlaProThrValProProValThrThrThrTh 741

Qy       1191 CAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTC 1250

Db        741 rProThrThrThrThrSerThrThrThrThrThrThrThrThrThrThrThrGlnTh 761

Qy       1251 TTCCGCGTCA------------TCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTC 1298

Db        761 rThrProThrThrProValValThrThrProSerThrIleThrProThrThrArg----- 779

Qy       1299 AATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTC 1358

Db        780 ----ProValThrGlnProAlaSerGlnProAlaThrGlnProAlaThrThrProThrTh 798

Qy       1359 GACG 1362

Db        798 rThr 799

RESULT 6
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
```

J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33532; MUID:89197956; PMID:2703501
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481; PMID:1985113
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t
A;Reference number: PQ0328; MUID:92198477; PMID:1550588
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Alignment Scores:

Pred. No.:	1.74e-05	Length:	3020
Score:	229.00	Matches:	162
Percent Similarity:	29.79%	Conservative:	65
Best Local Similarity:	21.26%	Mismatches:	251
Query Match:	5.65%	Indels:	284
DB:	2	Gaps:	36

US-09-914-958B-35 (1-2177) x A43932 (1-3020)

```
Qy       124 TGGCTCCGAGCGCTGACCCCGGCATGTCCA---------------------------- 153
             :::|||             |||   ||||||
Db      1159 TyrLeuGluGlyCysTyrProArgCysProLysAspArgProIleTyrGluGluAspLeu 1178

Qy       154 --------------------------GGATGTTAC---------------------- 162
                                        |||||||||
Db      1179 LysLysCysValThrAlaAspLysCysGlyCysTyrValGluAspThrHisTyrProPro 1198

Qy       163 ----------------------------CGTTCCTGCTGCTGCTCT------------ 180
                                          :::||||||   |||:::
Db      1199 GlyAlaSerValProThrGluGluThrCysLysSerCysValCysThrAsnSerSerGln 1218

Qy       181 ------------------------------------------------------GGTTTCTGC 189
                                                                    ||||||
Db      1219 ValValCysArgProGluGluGlyLysIleLeuAsnGlnThrGlnAspGlyAlaPheCys 1238

Qy       190 CCA---TCACTGAGGGGTCCCAGCGGG--------------CTGAACCCATGTTCACTG 231
                 :::  ||||||:::|||              |||   ||||||:::
Db      1239 TyrTrpGluIleCysGlyProAsnGlyThrValGluLysHisPheAsnIleCysSerIle 1258

Qy       232 ------------------------------------------------CAGTCACCAACT 243
                                                               :::||||||
Db      1259 ThrThrArgProSerThrLeuThrThrPheThrThrIleThrLeuProThrThrProThr 1278

Qy       244 CAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTA 303
             |||          |||   |||      |||   |||||||||:::          |||
Db      1279 SerPheThrThrThrThrThrThrThrProThrSerSerThrVal---------Leu 1295

Qy       304 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGG--------------------- 342
                         :::   |||  |||  |||
Db      1296 SerThrThrProLysLeuCysCysLeuTrpSerAspTrpIleAsnGluAspHisProSer 1315

Qy       342 ------------------------------------------------------------ 342

Db      1316 SerGlySerAspAspGlyAspArgGluProPheAspGlyValCysGlyAlaProGluAsp 1335

Qy       343 ---------CGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGA--- 390
                     |||        ||||||                  |||   :::
Db      1336 IleGluCysArgSerValLysAspProHisLeuSerLeuGluGlnHisGlyGlnLysVal 1355

Qy       391 ----------------------------------------------------AGCGGC--- 396
                                                                   :::|||
Db      1356 GlnCysAspValSerValGlyPheIleCysLysAsnGluAspGlnPheGlyAsnGlyPro 1375

Qy       397 ----------------------------------------TGGTGAACATCGCGGTCG 414
                                                     |||         :::
Db      1376 PheGlyLeuCysTyrAspTyrLysIleArgValAsnCysCysTrpProMetAspLysCys 1395

Qy       415 ATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCA 474
             ::::::          |||  |||                  ||||||   |||
Db      1396 IleThrThrProSerProProThrThr----------------ThrPro------Ser 1407

Qy       475 CAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATG 534
               |||   |||:::|||  |||          :::|||   |||  |||
Db      1408 ProProProThrThrThrThrThrLeuProProThrThrThrProSerProProThrThr 1427

Qy       535 CCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGG 594
               :::         |||      ||||||:::      |||      ||||||
Db      1428 ThrThrThrThrProPro------ProThrThrThrProSerProProIleThr------ 1443

Qy       595 AAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC 654
                                     :::|||   |||:::      |||   :::   |||
Db      1444 -----------------ThrThrThrThrProLeu------ProThrThrThrProSer 1455
```

```
Qy       655 GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATT 714
Db      1456 Pro------ProIleSerThrThrThrThrProProProThrThrThrProSerProPro 1473

Qy       715 ACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACC 774
Db      1474 ThrThrThr----------------ProSer--------------ProProThrThr 1482

Qy       775 TCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATA 834
Db      1483 ThrProSerPro------------------------------------------------ 1486

Qy       835 CAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCT 894
Db      1487 ------ProThrThrThrThrThrProProProThrThrThrPro---------Ser 1501

Qy       895 GCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGG 954
Db      1502 ProProMetThrThrProIleThrProProAlaSerThrThrThrLeuProProThrThr 1521

Qy       955 ACG---CTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCC 1011
Db      1522 ThrProSerProProThrThrThrThrThrThrProProProThrThrThrProSerPro 1541

Qy      1012 TGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCC 1071
Db      1542 ---ProThrThrThrProIleThrProProThrSerThr------ThrThrLeuProPro 1558

Qy      1072 ACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCAC 1131
Db      1559 ThrThrThr---------------ProSerProProProThrThrThrThrThrProPro 1573

Qy      1132 CCA---AGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATG 1188
Db      1574 ProThrThrThrProSerProProThrThrThrThrProSerProProThrIleThrThr 1593

Qy      1189 ACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCC 1248
Db      1594 Thr--------------------ThrProProProThrThrThrProSerProProThr 1606

Qy      1249 TCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCG 1308
Db      1607 ThrThrThrThrThrProProProThrThrThrProSerProProThrThrThrProIle 1626

Qy      1309 ACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACG 1368
Db      1627 ThrPro-----------------------------------ProThrSerThr------ 1632

Qy      1369 GGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCC 1428
Db      1633 ---------ThrThrLeuProProThrThrThrPro------SerProProProThrThr 1647

Qy      1429 GGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGG 1488
Db      1648 ThrThrThrProProProThrThrThr---------------------------------- 1656

Qy      1489 CCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGA 1548
Db      1657 ProSerPro---------ProThrThrThrThrPro---------SerProProIleThr 1670

Qy      1549 TCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGG 1608
Db      1671 ThrThrThrThrProProProThrThrThrProSerSerProIleThrThr--------- 1687

Qy      1609 GGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGA 1668
Db      1688 ------------------------------------------------------ThrPro 1689

Qy      1669 ACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACA 1728
Db      1690 SerProProThrThrThrMetThrThrProSerProThrThrThrProSerSerProIle 1709

Qy      1729 CACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCC 1788
```

```
Db      1710 ThrThrThrThrThrProSerSerThrThrThrProSerProProProThrThrMetThr 1729

Qy      1789 ATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACA------ 1842
Db      1730 ThrPro---SerProThrThrThrProSerProProThrThrThrMetThrThrLeuPro 1748

Qy      1843 ---------------AGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCA 1887
Db      1749 ProThrThrThrSerSerProLeuThrThrThrProLeuProProSerIleThrProPro 1768

Qy      1888 ACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCG------TGG 1941
Db      1769 ThrPheSerProPheSerThrThrThrProThrThrProCysValProLeuCysAsnTrp 1788

Qy      1942 GCTGGT 1947
Db      1789 ThrGly 1790
```

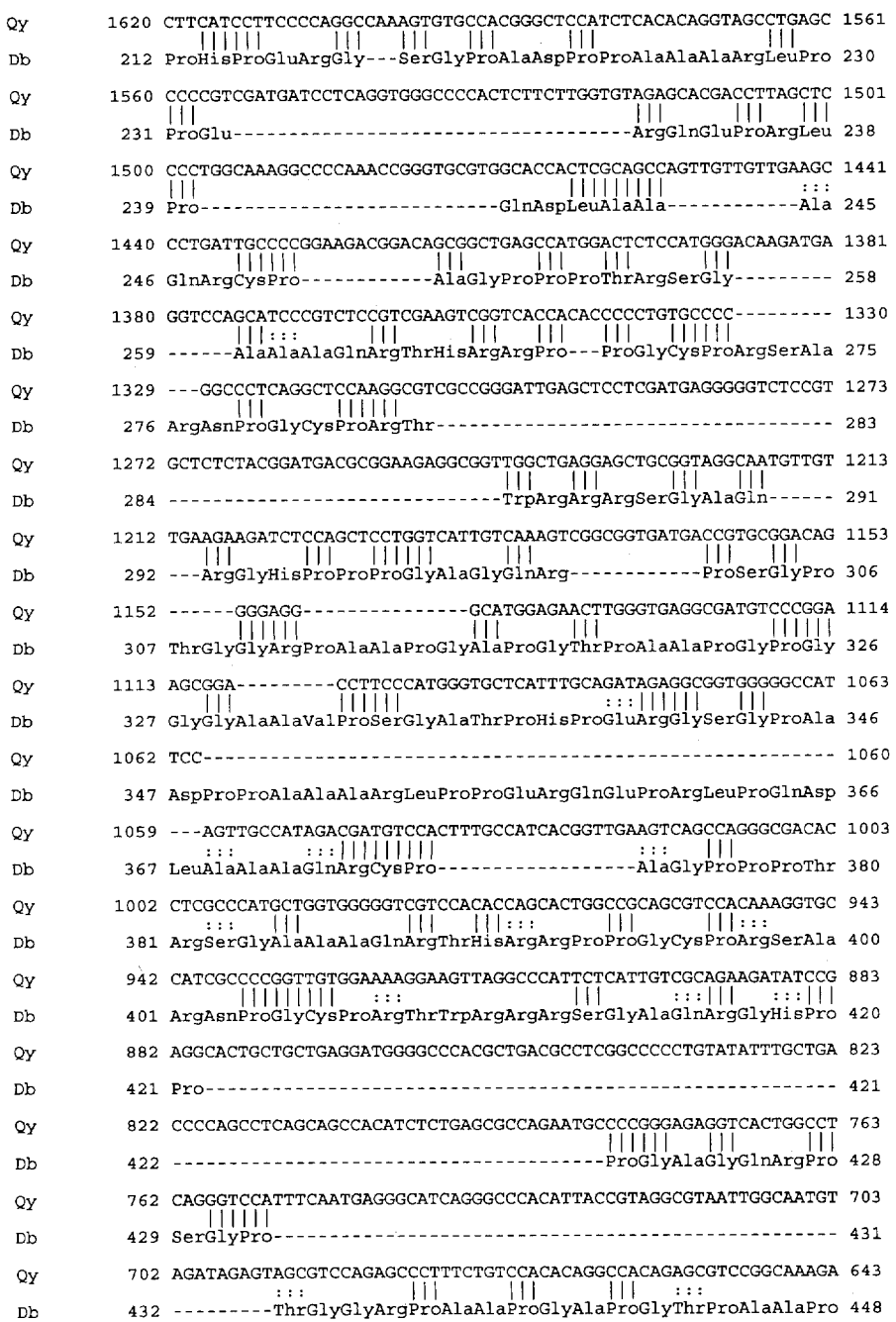
RESULT 7
QQBE3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C;Accession: A03742
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A;Reference number: A93065; MUID:85035713; PMID:6092825
A;Accession: A03742
A;Molecule type: DNA
A;Residues: 1-660 <BAN>
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; Ha
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667; PMID:6087149
A;Contents: annotation; protein coding region
C;Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C;Superfamily: human herpesvirus 4 BHLF1 protein

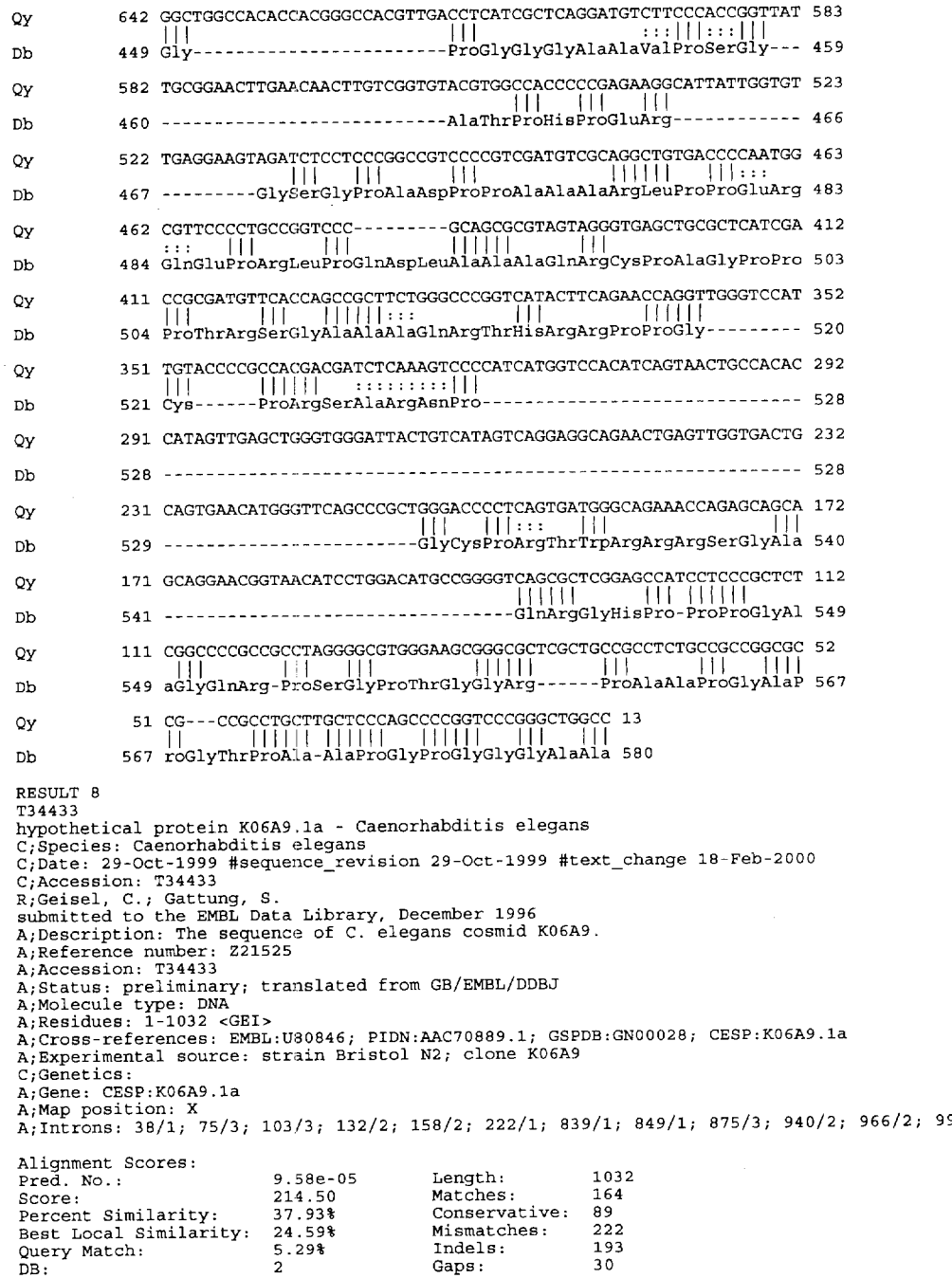
Alignment Scores:

Pred. No.:	2.64e-05	Length:	660
Score:	223.50	Matches:	162
Percent Similarity:	27.30%	Conservative:	28
Best Local Similarity:	23.28%	Mismatches:	218
Query Match:	5.41%	Indels:	288
DB:	1	Gaps:	34

US-09-914-958B-35 (1-2177) x QQBE3 (1-660)

```
Qy      1965 TCTTCAACACAGGGCTCCACCAGCCCACGCAGGCTGTGCCATCCT------------CGT 1918
Db       130 SerAlaGlyLeuGlySerArgGlyProArgPro------HisProAlaPheGlnValGln 147

Qy      1917 TGG---GCTCGTAGCCCCGACTGCACTTCTTGTTGGTCCGGCACCTGTAGCTTCCATAGG 1861
Db       148 TrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArg 167

Qy      1860 TGTTGACACATACGGGCTTGTCTCGAGGGCACACGAATGGGAACTGGATGCATTCATTGG 1801
Db       167 ------------------------------------------------------------ 167

Qy      1800 TGTCCATGCAATGGCCATTTTCCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTG 1741
Db       168 ----------------------------------------GlyHisProProProGlyAla 174

Qy      1740 GGTCCTGAAGTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCC 1681
Db       175 GlyGlnArgProSerGlyProThrGlyGlyArg------ProAlaAlaProGlyAlaPro 192

Qy      1680 CGCTGGCCACGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGG 1621
Db       193 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaVal---ProSerGlyAlaThr 211
```





RESULT 8
T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34433
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

Alignment Scores:

Pred. No.:	9.58e-05	Length:	1032
Score:	214.50	Matches:	164
Percent Similarity:	37.93%	Conservative:	89
Best Local Similarity:	24.59%	Mismatches:	222
Query Match:	5.29%	Indels:	193
DB:	2	Gaps:	30

US-09-914-958B-35 (1-2177) x T34433 (1-1032)

```
Qy       146 CATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTC 205
Db       219 AsnIleThrAlaAlaThrSerProSerThrAlaValThrSer-ProSerSerLeuGlyTh 238

Qy       206 CCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGA 265
Db       238 rSer---SerSerProLeuProSerSerIleSerThrSerAlaLeuProIleAlaSerSe 257

Qy       266 CAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGA 325
Db       257 rSerAlaSerSerSerProSerAlaAlaSerSerThrThrProVal-ValLeuSerSerS 277

Qy       326 CTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGC 385
Db       277 erThrIleGlnSerSerSerGlyThrPhePro---------------------------- 287

Qy       386 CCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCG 445
Db       288 --------------------------SerSerValAlaSerSerProSerThr---ValG 298

Qy       446 GGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGA 505
Db       298 lySerThrSerGlyAlaAlaSerSerSerTyrAlaThrValSerThrIleAlaGlyS 318

Qy       506 GGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGG------TGGCCACGTACAC 559
Db       318 erThrGlySerThrIleThrProValProGlySerSerSerThrIleGlySerSerThrP 338

Qy       560 CGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTG-------AGCGATG 612
Db       338 roSerAlaSerSerSerSerSerGlyThrMetSerThrIleSerGlySerThrGlySerT 358

Qy       613 AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCT---GTGGCCTGTG 669
Db       358 hrValThrValValProGlySerSerSerThrPheAlaSerSerThrProIleAlaSerS 378

Qy       670 TGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCC---------AATTACGCCT 720
Db       378 erSerSerProGlySer---------ThrValThrValAlaProGlySerSerSerThrT 395

Qy       721 ACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCC 780
Db       395 yrGlySer-----------------------SerThrProSerAlaSerSerSerSerS 407

Qy       781 GGGGCATTCTGGCGCTCAGA---------------GATGTGGCTGCTGAGGCTGGGGTCA 825
Db       407 erGlyThrMetSerThrAsnSerGlySerThrGlySerThrValThrValAlaProValS 427

Qy       826 GCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGG 885
Db       427 erSerSerThrPheGly---------SerSerThrProIleAlaSerSerSerSerSerG 444

Qy       886 ATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCA 945
Db       444 ly-SerThrValThrValValSerGly------SerSerSerThrTyrGlySerSerThr 461

Qy       946 CCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTG 1005
Db       462 ProSer----------------------------AlaSerSerSerSerAlaGlyThrAla 472

Qy      1006 TCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATG 1065
Db       473 Ser--------------------------------ThrIleSerGlySerThrGly--- 480

Qy      1066 GCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCG 1125
Db       481 ---SerThrAlaThrIle---------ValProGlySerSerSerSerValGlySerSer 496

Qy      1126 CCTCACCCAAGTTCTCCATGCCCTCCC------CTGTCCGCACGGTCATCACCGCCGACT 1179
Db       497 ThrGlnSerAlaSerProSerSerProGlyThrMetSerThrValSerGlyProThrGly 516

Qy      1180 TTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAG 1239
Db       517 SerThrValThrValValProGlySerSerThrSerProAlaProSerSerSerProAsn 536

Qy      1240 CCAACCGCCTCTTCCGCGTCATCC---------------------------------- 1263
Db       537 ProSerSerSerProAlaSerThrGlySerThrIleThrIleSerGlySerSerSerIle 556

Qy      1263 ------------------------------------------------------------ 1263
Db       557 IleValSerThrValSerGlySerThrValSerGlySerThrGlyThrSerGlnSerThr 576

Qy      1264 ---GTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGC 1320
Db       577 LeuAlaSerSerThrAlaThrProGlySerSerSerThrValProSerSerSer---Ser 595

Qy      1321 CTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACC 1380
Db       596 ProGlnProSerSerGlnSerProAlaProAsnThrGlySerThrThrProSerGlnThr 615

Qy      1381 TCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCC------------ 1428
Db       616 SerSer---------GlnSerProSerProSerMetAsnProSerSerSerThrProThr 632

Qy      1429 ----------------------------GGGGCAATCAGGGCTTCAACAACA------ 1452
Db       633 GlySerSerGlnSerThrIleThrProGluGlySerThrAlaSerSerProThrGlySer 652

Qy      1453 ACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCG 1512
Db       653 ThrGlySerThrPheSerValAlaThr--------------------GluValThrSer 665

Qy      1513 TGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACC 1572
Db       666 GlnSerThrVal------------ProSer---GlySerSerLeuGlyThrGlnSerThr 680

Qy      1573 TGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGG 1632
Db       681 AsnSerSerProSerPro------------------SerSerLeuSerPro--------- 691

Qy      1633 AGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACT 1692
Db       691 ------------------------------------------------------------ 691

Qy      1693 CAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGG 1752
Db       692 ---------------SerThrSerGlyMetSerThr---LeuThrSerGluProSerPro 705

Qy      1753 AGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCA 1812
Db       706 SerSerThrGlnSerSerGlyAlaGlnSerThrLeuThrThrProSerProAsnProSer 725

Qy      1813 TCCAGTTCCCATTCGTGTGCCCTCGAGACAAGC--------------------------- 1845
Db       726 GlnSerThrSerSerLeuGluSerSerThrSerGlyAlaThrThrSerSerGlySerAla 745

Qy      1846 ---CCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTC 1902
Db       746 GlyThrThrMetThrSerProSerGlnSerSerSerValGly---------SerSerGln 762

Qy      1903 GGGGCTACGAGCCCA 1917
Db       763 GlySerThrSerPro 767
```

RESULT 9
VGBEX1
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C;Species: equine herpesvirus 1
A;Note: host Equus caballus (domestic horse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C;Accession: H36802
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.

submitted to GenBank, March 1992
A;Description: The DNA sequence of equine herpesvirus-1.
A;Reference number: A36805
A;Accession: H36802
A;Molecule type: DNA
A;Residues: 1-797 <TEL>
A;Cross-references: GB:M86664; NID:g330791; PIDN:AAB02506.1; PID:g330862
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566; PMID:1318606
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-797/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	0.000104	Length:	797
Score:	213.50	Matches:	113
Percent Similarity:	35.70%	Conservative:	68
Best Local Similarity:	22.29%	Mismatches:	197
Query Match:	5.26%	Indels:	129
DB:	1	Gaps:	15

US-09-914-958B-35 (1-2177) x VGBEX1 (1-797)

```
Qy        403 ACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACG 462
Db         23 ThrThrThrThrGluThrThrThrSerSerSerSerThrSerGlySerGlyGlnSerThr 42

Qy        463 CCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCA 522
Db         43 SerSerGlyThrThrAsnSerSerSerSerProThrThrSerProProThrThrSerSer 62

Qy        523 ACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCA 582
Db         63 SerPro------ProThrSerThrHisThrSerSerProSerSerThrSerThrGlnSer 80

Qy        583 ATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCC 642
Db         81 SerSerThrAlaAlaThrSerSerSerAlaProSerThr----------------Ala 94

Qy        643 TCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCT 702
Db         95 SerSerThrThrSerIle---ProThrSerThrSerThrGluThrThrThrThrPro 113

Qy        703 ACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTG 762
Db        114 ThrAlaSerThrThrThrProThrThrThrThrAlaAlaProThrThrAlaAlaThrThr 133

Qy        763 AGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGG 822
Db        134 ThrAlaValThrThr------------------------------------AlaAla 140

Qy        823 TCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
Db        141 SerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSerThrProThrThrThr 160

Qy        883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
Db        161 ThrProThrSerThrThrThrThrThrAlaThrThrThrValProThrThrAlaSerThr 180

Qy        943 GCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAG 1002
Db        181 ThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrThrAlaAlaThr 200

Qy       1003 GTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGA 1062
Db        201 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 220

Qy       1063 ATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACA 1122
Db        221 ThrAlaAlaThrThrSer--------------------------------SerAlaThr 229

Qy       1123 TCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTG 1182
Db        230 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 249

Qy       1183 ACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCA 1242
Db        250 Thr------------------AlaAlaThrThrThrAlaAlaThrThrThrGlySerPro 263

Qy       1243 ACCGCCTCTTCCGCGTCATCCGTA---GAGAGCACGGAGACCCCCTCATCGAGGAGCTCA 1299
Db        264 ThrSerGlySerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSerThrAla 283

Qy       1300 ATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCG 1359
Db        284 ThrSerAlaThrProThrSer--------------------------------ThrSer 292

Qy       1360 ACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGT 1419
Db        293 ThrSerAlaAlaAlaThrThrSer----------------------------ThrProThr 303

Qy       1420 CCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCC 1479
Db        304 ProThrSerAlaAlaThrSerAlaGluSerThrThr---------------GluAlaPro 318

Qy       1480 GGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCC 1539
Db        319 --------------------------------ThrSerThr---------------Pro 322

Qy       1540 ACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACT 1599
Db        323 ThrThrAspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThr 342

Qy       1600 TTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGG 1659
Db        343 ThrValSerAla---------------------------------------------- 346

Qy       1660 TGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGG 1719
Db        347 ---------------------------------------------SerThrThrSerAla 351

Qy       1720 ATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGG 1779
Db        352 ThrThrThrAlaPheThrThrGluSerHisThrSer------------------------ 363

Qy       1780 AAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAG 1839
Db        364 --------------------ProAspSerSerThrGlySerThrSerThrAlaGluPro 376

Qy       1840 ACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCA 1899
Db        377 SerSerThrPheThrLeuThrProSerThrAlaThrProSer------ThrAspGlnPhe 394

Qy       1900 GTCGGGGCTACGAGCCCAACG 1920
Db        395 ThrGlySerSerAlaSerThr 401
```

RESULT 10
T45463
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 17-Mar-2000
C;Accession: T45463
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J. Equine Sci. 7, 79-87, 1996

A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equin
A;Reference number: Z22973
A;Accession: T45463
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-867 <KIR>
A;Cross-references: EMBL:D88734; PIDN:BAA20038.1
A;Experimental source: isolate 3F clone; strain BK343
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo

Alignment Scores:

Pred. No.:	0.00018	Length:	867
Score:	209.50	Matches:	127
Percent Similarity:	33.07%	Conservative:	82
Best Local Similarity:	20.09%	Mismatches:	284
Query Match:	5.17%	Indels:	139
DB:	2	Gaps:	15

US-09-914-958B-35 (1-2177) x T45463 (1-867)

```
Qy        25 ACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGCGCCCGC 84
Db        36 SerGlySerGlyGlnSerThrSerSerGlyThrThrAsnSerSerSerSerProThrThr 55

Qy        85 TTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTGACCCCG 144
Db        56 SerProProThrThrSerSerSerProProThrSerThrHisThrSerSerProSerSer 75

Qy       145 GCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGT 204
Db        76 ThrSerThrGlnSerSerSerThrAlaAlaThrSerSerSerAlaProSerThrAlaSer 95

Qy       205 CCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATG 264
Db        96 SerThrThrSerIleProThrSerThrSerThrGluThrThrThrThrProThrAla 115

Qy       265 ACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGG 324
Db       116 SerThrThrThrProThrThrThr---------------------------------- 123

Qy       325 ACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGG 384
Db       124 ------------------------ThrAlaAlaProThr--------------ThrAla 130

Qy       385 CCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGC 444
Db       131 AlaThrThrThrAlaValThrThrAlaAlaSerThrSerAlaGluThrThrThrAlaThr 150

Qy       445 GGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGG 504
Db       151 AlaThrAlaThrSerThrProThrThrThrThrProThrSerThrThrThrThrThr--- 169

Qy       505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
Db       170 ------AlaThrThrThrValProThrThrAlaSerThrThrThrAspThrThrAla 187

Qy       565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
Db       188 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAla 207

Qy       625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
Db       208 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAla 227

Qy       685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
Db       228 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr------------------ 241

Qy       745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
Db       241 ------------------------------------------------------------ 241

Qy       805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
Db       242 ------------------AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 255

Qy       865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
Db       256 SerSerAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 275

Qy       925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
Db       276 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 295

Qy       985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
Db       296 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 315

Qy      1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
Db       316 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrGlySerProThrSer 335

Qy      1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
Db       336 GlySerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSer---ThrAlaThr 354

Qy      1165 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 1224
Db       355 SerAlaThrProThrSerThr------------------SerThrSerAlaAlaAlaThr 368

Qy      1225 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 1284
Db       369 ThrSerThrProThrProThrSerAlaAlaThrSerAlaGluSerThrThrGluAlaPro 388

Qy      1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
Db       389 ThrSer------------------ThrPro------------------------------ 392

Qy      1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
Db       393 ------------ThrThrAspThr--------------ThrThrProSerGluAlaThr 403

Qy      1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
Db       404 ThrAlaThrThrSerProGluSerThrThrValSerAlaSerThrThrSer--------- 420

Qy      1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
Db       421 --------------------------------AlaThrThrThrAlaPheThrThrGlu 429

Qy      1525 AGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
Db       430 SerHisThrSerPro------AspSerSerThrGlySerThrSerThrAlaGluProSer 447

Qy      1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
Db       448 SerThrPhe---ThrLeuThrProSerThrAlaThrProSerThrAspGlnPheThrGly 466

Qy      1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
Db       467 SerSerAla---------------------------------------------------- 469

Qy      1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
Db       470 ---SerThrGluSerAspSerThrAspSerSerThrValProThrThrGlyThrGlu--- 487

Qy      1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
Db       488 --------------------------SerIleThrGluSerSerSerThrThrGluAla 498

Qy      1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
Db       499 SerThrAsnLeuGlySerSerThrTyrGluSerThrGluAlaLeuGluThrProAspGly 518
```

```
Qy       1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
Db        519 AsnThrThrSer------GlyAsnThrThrProSer 528

RESULT 11
T45462
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jun-2000
C;Accession: T45462
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J. Equine Sci. 7, 79-87, 1996
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equin
A;Reference number: Z22973
A;Accession: T45462
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-866 <KIR>
A;Cross-references: EMBL:D88733; PIDN:BAA20037.1
A;Experimental source: strain HH1
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo
F;558-866/Domain: equine herpesvirus 1 glycoprotein homology <EHG>

Alignment Scores:
Pred. No.:                0.000267      Length:          866
Score:                    206.50        Matches:         125
Percent Similarity:       34.31%        Conservative:    74
Best Local Similarity:    21.55%        Mismatches:      233
Query Match:              5.09%         Indels:          148
DB:                       2             Gaps:            18

US-09-914-958B-35 (1-2177) x T45462 (1-866)

Qy        195 ACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCA--------- 245
Db         35 ThrSerGlySerGlyGlnSerThrSerSerGlyThrThrAsnSerSerSerSerProThr 54

Qy        246 GTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACT 305
Db         55 ThrSerProProThrThrSerSerSer-ProProThrSerThr----------------- 68

Qy        306 GATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTG 365
Db         69 ----------------------------------------HisThrSerSerPro----- 73

Qy        366 GTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGC 425
Db         74 -------SerSerThrSerThrGlnSerSerSer---ThrAlaAlaThrSerSerSerAl 90

Qy        426 TCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGAC 485
Db         90 aProSerThrAlaSerSerThrThrSerIleProThrSerThrSerThrGluThrThrTh 110

Qy        486 ATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGG 545
Db        110 rThrThrProThrAlaSerThrThrThrProThrThrThrThrAlaAlaProThrThrAl 130

Qy        546 GTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTG 605
Db        130 a------AlaThrThrThrAlaValThrThrAlaAlaSerThrAlaAlaSerThrSer-- 147

Qy        606 AGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCC 665
Db        148 ----AlaGluThrThrThrAlaThrAlaThrAlaThrSerThrProThrThrThrThrPr 166

Qy        666 TGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGT 725
Db        166 o------ThrSerThrThrThrThrThrAlaThrThrThrValProThrThrAlaSerTh 184

Qy        726 AATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGC 785

Db        184 rThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrTh 204

Qy        786 ATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGA 845

Db        204 r----------------------------------------------------ThrAl 206

Qy        846 GGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAG 905

Db        206 aAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAl 226

Qy        906 AATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCC 965

Db        226 aAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrSerSe 246

Qy        966 AGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAAC 1025

Db        246 rAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAl 266

Qy       1026 CGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTG 1085

Db        266 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAl 286

Qy       1086 CAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATG 1145

Db        286 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAl 306

Qy       1146 CCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATC 1205

Db        306 aThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr-----------------Al 320

Qy       1206 TTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGT 1265

Db        320 aAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGlySerThrSerThrTh 340

Qy       1266 A---GAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCT 1322

Db        340 rGlyAlaSerThrSerThrProSerAlaSerThrAlaThrSerAlaThrProThrSer-- 359

Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382

Db        360 ------------------------------ThrSerThrSerAlaAlaAlaThrThrSe 369

Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGC 1442

Db        369 r--------------------------ThrProThrProThrSerAlaAlaThrSerAl 380

Qy       1443 TTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGA 1502

Db        380 aGluSerThrThr---------------GluAlaPro---------------------- 387

Qy       1503 GCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGC 1562

Db        388 ----------ThrSerThr---------------ProThrThrAspThrThrThrProSe 399

Qy       1563 TCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCC 1622

Db        399 rGluAlaThrThrAlaThrThrSerProGluSerThrThrValSerAla----------- 415

Qy       1623 AGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGG 1682

Db        415 ------------------------------------------------------------ 415

Qy       1683 GAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCA 1742

Db        416 ----------------------SerThrThrSerAlaThrThrThrAlaPheThrThrGl 428

Qy       1743 GCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACC 1802

Db        428 uSerHisThrSer-------------------------------------------Pr 433

Qy       1803 AATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACC 1862

Db        433 oAspSerSerThrGlySerThrSerThrAlaGluProSerSerThrPheThrLeuThrPr 453
```

```
Qy       1863 TATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920
Db        453 oSerThrAlaThrProSer------ThrAspGlnPheThrGlySerSerAlaSerThr 470

RESULT 12
T48814
hypothetical protein 15E6.220 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C;Accession: T48814
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24541
A;Accession: T48814
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1952 <SCH>
A;Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.220
A;Experimental source: cosmid contig 15E6; strain 74
C;Genetics:
A;Gene: NCSP:15E6.220
A;Map position: 2
A;Introns: 281/3

Alignment Scores:
Pred. No.:                0.000362        Length:          1952
Score:                    205.50          Matches:         162
Percent Similarity:       33.82%          Conservative:    70
Best Local Similarity:    23.62%          Mismatches:      267
Query Match:              5.07%           Indels:          187
DB:                       2               Gaps:            32

US-09-914-958B-35 (1-2177) x T48814 (1-1952)

Qy        418 AGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAG 477
Db       1251 AlaAlaGlnArgIleAlaThrProSerAlaThrGlnArgAlaAsnProProGlyGlyAsn 1270

Qy        478 CCTGCGACA------TCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATA 531
Db       1271 ProSerAlaValProArgThrIleThrAlaAlaAlaProAlaProThrProAlaProLeu 1290

Qy        532 ATGCCTTCT------------CGGGGGTGGCCACGTACACCGACAAGTTGT---TCAAGT 576
Db       1291 LeuProArgAlaAlaProIleSerGlyHisProIleAlaProSerThrValAsnAsnThr 1310

Qy        577 TCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGG 636
Db       1311 AsnAlaAlaMetGlyAlaArg--------------------------ProVal------ 1319

Qy        637 CCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACT 696
Db       1320 ProAlaAlaValProGlyProGlnIleGlyValHisSerGlyAlaAlaIleGlyValThr 1339

Qy        697 CTA---------TCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
Db       1340 AlaProValSerArgSerProProValThrHisGlyValVal------------------ 1353

Qy        748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
Db       1354 ---------ThrAlaAlaProThrThrThrPro------------ArgSerAsnLeuAla 1366

Qy        808 CTG-------------------------------------------------CTGAGGCTG 819
Db       1367 AlaSerValProAlaGlySerMetAlaGlnGlnThrHisSerGlyAlaProAlaGlySer 1386

Qy        820 GGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCA---GCAGCA 876
Db       1387 GlyAsnAlaSerIleSerArgAlaProProSerThrAlaAlaProThrAlaProThrSer 1406

Qy        877 GTGCCTCGGATATCTTCTGCGACAATGAGAATG-----------------------GGC 912
Db       1407 ValProProValSerSerThrValArgProMetSerTyrValProThrGlyProProGly 1426

Qy        913 CTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTG 972
Db       1427 LeuThrAlaProProThrAlaSerSerGlyAlaSerAlaGlyTyrAlaArgProAlaAsn 1446

Qy        973 GTGTGGACGACCCCC---------ACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCA 1023
Db       1447 AlaSerThrMetProAlaProThrSerGlyGlnAlaAlaMetThrSerAlaValProGln 1466

Qy       1024 ACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATC 1083
Db       1467 SerValProSerProArgProSerSerLeuThrThrGlyGlnMetGlyValProAlaIle 1486

Qy       1084 TGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCA 1143
Db       1487 AlaAla---AlaSerThrSerArgProAlaSerGlyValTyrAsnProProAlaSerSer 1505

Qy       1144 TGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGA 1203
Db       1506 LeuAlaProSerThrHisLysSerMetProSerAlaValProThrThrAlaSerGlyAla 1525

Qy       1204 TCTTCTTCAACAACA--------------------TTGCCTACCGCAGCTCCTCAG--- 1239
Db       1526 ValSerSerThrValSerSerLeuAlaAlaThrProLeuProProSerAlaProArgTyr 1545

Qy       1239 ------------------------------------------------------------ 1239
Db       1546 GlyProSerAsnAsnAlaThrProThrThrProAlaThrAlaIleProArgAlaProLeu 1565

Qy       1240 CCAACCGCCTCTTCCGCGTCATCC---GTAGAGAGCACGGAGACCCCCTCATCGAGGAGC 1296
Db       1566 ProAlaAlaSerSerValSerAlaProValThrGlyGlnProSerPheSerAlaProAla 1585

Qy       1297 TCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACT 1356
Db       1586 SerValProThrProProThrSer------AlaAlaSerGlnGlyAlaGln---ProLeu 1602

Qy       1357 TCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGC 1416
Db       1603 SerGlnProValGly------------ProAlaProAlaIleSerSerSerIleGlyAla 1618

Qy       1417 TGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCA 1476
Db       1619 ThrProAla--------------AlaSerIleProSerSerAlaProAlaAlaLeuAla 1633

Qy       1477 CCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGG 1536
Db       1634 ProValThrTyrProValProGlnGlnAlaSerAlaAlaAlaAlaArgLeuProValThr 1653

Qy       1537 CCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGT------------------GTG 1578
Db       1654 ProAlaProAlaAlaHisThrIleAlaGlnSerValAlaGlnProValAlaArgProVal 1673

Qy       1579 AGATGGAGCCCG---------------------------------------------- 1590
Db       1674 ThrGlnSerProValGlnSerValAlaGlnHisValThrSerSerGlnAlaSerThrThr 1693

Qy       1591 ---TGGCACACTTTGGCCTGGGGAAGGATGAAGCCA----------------------- 1623
Db       1694 AlaAlaHisProValAlaGlnSerValProArgProValAsnSerAsnProThrSerAla 1713

Qy       1624 ------------GCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACG 1671
Db       1714 AlaProValAlaAlaValGly------ThrAlaGlnValAlaSerAlaProThrValThr 1731

Qy       1672 TGGCCAGCG---GGGAGATGAACTCAGTGCTGGAGATCCCTCTACCCCCGGGATGAGGACA 1728
Db       1732 GlnProAlaProHisArgAlaLeuSerSerValSerGlnSerLeuProGlnSerValPro 1751

Qy       1729 CACTTCAGGACCCAG--------------CCCCAC----------------------- 1749
Db       1752 HisAlaAlaGlnGlnAlaAlaHisGlnThrProHisSerAlaSerArgProValProGln 1771
```

```
Qy       1750 ---------TGGAGTGTGGCCAAGGATTCTCCCAGC---------AGGAAAATGGCCATT 1791
Db       1772 SerValProGlnSerValProGlnAlaThrProGlnAlaValProArgProSerThrSer 1791

Qy       1792 GCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGC------ 1845
Db       1792 AlaLeuThrProThrAlaGlnProGlyProValSerProAlaValSerGlySerGlyVal 1811

Qy       1846 ----------------------------------CCGTATGTGTCAACACCTATGGAA 1869
Db       1812 ProAlaProSerAlaAlaGlnSerValAlaProAlaProValSerSerThrProValPro 1831

Qy       1870 GCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATG--- 1926
Db       1832 AlaAlaThrValAlaProAlaSerThrValAlaAlaAlaProThrProThrArgValThr 1851

Qy       1927 GCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGA 1986
Db       1852 AlaAlaProAlaAlaLeuSerAlaAla--------------------------ThrAsn 1862

Qy       1987 AGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
Db       1863 ProAlaProValProSerGlnProGlnHisGlnIleThrGlyGln------AlaProAla 1880

Qy       2047 GATCAGGAACACTTACCT 2064
Db       1881 GlnGlnGlnArgProPro 1886

RESULT 13
S35047
mucin JUL7 - human
C;Species: Homo sapiens (man)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S35047
R;Dufosse, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperat, V.; Laine, A.; van-Seuningen,
Biochem. J. 293, 329-337, 1993
A;Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternati
A;Reference number: S35047; MUID:93343858; PMID:7916618
A;Accession: S35047
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-543 <DUF>
A;Cross-references: EMBL:X74370; NID:g407081; PIDN:CAA52408.1; PID:g407082
A;Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residue 3
A;Note: the amino acid sequence from Fig. 2 is inconsistent with the nucleotide sequence

Alignment Scores:
Pred. No.:              0.000362        Length:         543
Score:                  203.50          Matches:        145
Percent Similarity:     36.14%          Conservative:   61
Best Local Similarity:  25.44%          Mismatches:     211
Query Match:            5.02%           Indels:         153
DB:                     2               Gaps:           31

US-09-914-958B-35 (1-2177) x S35047 (1-543)

Qy         78 CGCCCGCTTCCCACGCCCCTA---GGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCG--- 131
Db         18 ArgProCysProProProLeuArgGlyArgProGlySerSerThrGluLeuAlaProThr 37

Qy        132 --------------------------------AGCGCTGACCCCGGCATGTCCAGGATG 158
Db         38 IleSerHisTyrAspCysSerProLeuAspSerThrAla-ThrProSerThrProGlyTh 57

Qy        159 TTAC---------CGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCCCAG 209
Db         57 rAlaProProProLysValLeuThrSerProAlaThrThrProThrAlaThrSerSerLy 77

Qy        210 CGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGT 269
Db         77 sAlaThrSerSerSerSerProArgThrAlaThrThrLeuProValLeuThrSerThrAl 97

Qy        270 AATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTT 329
Db         97 aThrLysSerThrAlaThrSerValThrProIleProSerSerThr---LeuGlyThrTh 116

Qy        330 GAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAG 389
Db        116 rGlyThrLeuProGluGlnThrThrThrProVal-------------------------- 127

Qy        390 AAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT---ACTACGCGCTGCGG 446
Db        128 ------------------AlaThrMetSerThrIleHisProSerSerThrProGluTh 141

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db        141 rThrHisThrSerThrValLeuThrThrLysAlaThrThrThrArgAlaThrSerSerTh 161

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db        161 rSerThrProSerSerThr---------ProGlyThrThrTrpIleLeuThrGluLeuTh 178

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db        178 rThrAlaAlaThrThrThrAlaGly--------------------------ThrGlyPr 189

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db        189 oThrAlaThrProSerSerThrProGlyThrThrTrp---IleLeuThrGluLeuThrTh 208

Qy        687 GGACGCTACT--------------------CTATCTACATTGCCAATTACGCCTACGGT 725
Db        208 rThrAlaThrThrThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyTh 228

Qy        726 AATGTGGGCCCTGATG---CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGG 782
Db        228 rThrTrpIleLeuThrGluProSer---------ThrThrAlaThrValThrValProTh 245

Qy        783 GGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGC 842
Db        245 rGly----------------------------------SerThrAlaThrAlaSerSerTh 254

Qy        843 CGAGGCGTCAGCGTGGGCCCCA---------TCCTCAGCAGCAGTGCCTCGGATATCTTC 893
Db        254 rGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThrValThrSe 274

Qy        894 TGCG--------------------------ACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db        274 rSerLysProLeuLeuLeuSerProGlyLeuGlnLeuProSerAlaLeuArgSerThrAl 294

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACG---------CTGCGGCCAGTGCTGGTGTG 977
Db        294 aThrAsnThrHisSerTyrGlnLeuTyrSerHisProLeuLeuLeuProGlyThrThrTr 314

Qy        978 G------------------ACGACCCCCACC---AGCATGGGCGAGGTGTCGCCCTGGCT 1016
Db        314 pThrArgLeuSerGlnThrThrThrProThrAlaThrMetSerThrAlaThrProSerSe 334

Qy       1017 G------------------ACTTCAACCGTGATG---------GCAAAGTGGACATCGTC 1049
Db        334 rThrProGluThrValHisThrSerThrValLeuThrAlaThrAlaThrThrThrGlyAl 354

Qy       1050 TATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTC 1109
Db        354 aProAla-Leu---TrpProProPro--------------------------------P 362

Qy       1110 CGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATC 1169
Db        362 roProProGlnGluGlnLeuThrLeuProLysCysArgLeuProGlnProArgLeuHisS 382

Qy       1170 ACCGCCGACTTTGACAATGACCAGGAGCTGGA----------------GATCTTCTTC 1211
Db        382 erHisProLeuLeuGln---ProArgAspGlyThrHisAlaSerSerValAspGlnHisA 401

Qy       1212 AACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAG 1271
```

```
Db        401 snHisHis-ThrHisAsnHisHisThrHisAsnGlnTrpLeu------------------ 414

Qy       1272 CACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGG 1331
Db        415 HisGlyAspProLeu---------LeuHisProGlyAspHisProHisArgGlnSerAla 431

Qy       1332 --------------------GGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGG 1370
Db        432 AspHisHisHisHisAsnCysAlaThrGlySerMetAlaThr-ProSerSerSerThr-- 450

Qy       1371 ATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGG 1430
Db        451 ----GlnThrSerGlyThrPro---------ProSerLeuThrThrThrAlaThrThrIl 466

Qy       1431 GGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCC 1490
Db        466 eThrAlaThrGlySerThrThrAsnProSerSer---------ThrProGlyThrThrPr 483

Qy       1491 TTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATC 1550
Db        483 oIleProProValLeuThrSerThrAlaThrThr---------------------ProAl 496

Qy       1551 ATCGACGGGGGCTCAGGCTACC 1572
Db        496 aAlaThrSerSerLysAlaThr 503

RESULT 14
T21460
hypothetical protein ZK945.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21460
R;Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A;Reference number: Z19425
A;Accession: T21460
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-796 <WIL>
A;Cross-references: EMBL:Z48582; PIDN:CAA88469.1; GSPDB:GN00020; CESP:ZK945.10
A;Experimental source: clone F27E5
C;Genetics:
A;Gene: CESP:ZK945.10
A;Map position: 2
A;Introns: 62/3; 124/3; 169/1; 216/1; 262/3; 371/1; 581/3; 608/1

Alignment Scores:
Pred. No.:              0.000392        Length:         796
Score:                  203.50          Matches:        151
Percent Similarity:     36.97%          Conservative:   83
Best Local Similarity:  23.85%          Mismatches:     210
Query Match:            5.02%           Indels:         189
DB:                     2               Gaps:           28

US-09-914-958B-35 (1-2177) x T21460 (1-796)

Qy        190 CCATCACTGAGG---------GGTCCCAGCGGGCTGAACCCATGT------TCACTGCAG 234
Db        253 ProThrLeuArgArgMetLysArgAspAlaGlyAspAsnThrCysAspTyrThrIleGlu 272

Qy        235 TCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTG 294
Db        273 SerThrSer--------------ThrSerThrThrThrProThrThrThrThrValThr 287

Qy        295 TGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATG 354
Db        288 --------------------------SerThrValThrSerThr---------ThrThr 295

Qy        355 GACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCG 414
Db        296 ValProThrSerThrSerThrValThr--------------------ThrAlaMetSer 308

Qy        415 ATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCA 474
Db        309 ThrSerThrSerThrProSerThr-----------------SerThrThrIleGluSer 322

Qy        475 CAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATG 534
Db        323 ThrSerThrThrPheThrSerThrAlaSer---ThrSerThrSerSerThr--------- 338

Qy        535 CCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGG 594
Db        339 --------------------SerThrThrGlnGlnSerSerSerThrIleThrSerSer 351

Qy        595 AAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC 654
Db        352 ProSerSer------------------------------------------------- 354

Qy        655 GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACA---TTGCCA 711
Db        355 ---------------------------------ThrThrLeuSerThrSerIlePro 362

Qy        712 ATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTG 771
Db        363 ThrThrThrThr------------------ProGluIleThrSerThrLeuSerSerLeu 376

Qy        772 ACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAAT 831
Db        376 ------------------------------------------------------------ 376

Qy        832 ATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCT 891
Db        377 ----------------------------ProAspAsnAlaIleCysSerTyrLeuAsp 386

Qy        892 TCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTG 951
Db        387 GluThrThrThrSer---------ThrThrPheThrThrThr------------------ 397

Qy        952 TGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCC 1011
Db        398 ----------------------MetLeuThrSerThrThrThrGluGluProSerThr 409

Qy       1012 TGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCC 1071
Db        410 SerThrThrThrThrGluValThrSer---ThrSerSerThrValThrThrThrGluPro 428

Qy       1072 ACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCAC 1131
Db        429 ThrThrThrLeuThrThr--------------SerThrAlaSerThrSerThrThrGlu 443

Qy       1132 CCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACC 1191
Db        444 ProSerThrSerThrValThrThrSerProSerThrSerProValThrSerThrValThr 463

Qy       1192 AGGAGCTGGAGATCTTCTTCAACA---ACATTGCCTACCGCAGCTCCTCAGCCAACCGCC 1248
Db        464 SerSerSerSerSerSerThrThrValThrThrProThrSerThrGluSerThrSerThr 483

Qy       1249 TCTTCCGCGTCATCCGTA---GAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCG 1305
Db        484 SerProSerSerThrValThrThrSerThrThrAlaProSerThrSerThrThrGlyPro 503

Qy       1306 GCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAG 1365
Db        504 SerSerSerSerSerThr--------------------------ProSerSerThrAla 514

Qy       1366 ACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCT 1425
Db        515 SerSerSerValSerSerThrAlaSerSerThrGlnSerSerThrSerThrGlnGlnSer 534

Qy       1426 TCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTG 1485
Db        535 SerThrThrThrLysSerGluThrThrThr-SerSerAspGly---ThrAsnProAspPh 553

Qy       1486 GGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT------------ 1532
```

```
Db        553 e-TyrPheValGluLysAlaThrThrThrPheTyrAspSerThrSerValAsnLeuThrL 573

Qy       1533 --GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCG 1590
Db        573 euAsnSerGlyLeuGlyIleIle-----------GlyTyrGlnThrSerIleGluCysT 589

Qy       1591 TGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATG 1650
Db        589 hrSer-----------------ProThrSerSerAsnTyrValSerThrThrLysAspG 603

Qy       1651 GCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCT 1710
Db        603 lyAlaCysPheThrLysSerValSerMetProArgLeuGlyGlyThr-TyrProAlaSer 622

Qy       1711 ACC---------CCCGGGATGAGGACACACTTCAGGACC---------------------- 1740
Db        623 ThrPheValGlyProGlyAsnTyrThrPheArgAlaThrMetThrThrAspAspLysLys 642

Qy       1741 ----------------------------CAGCCCCACTGGAGTGTGGCCAAGGATTCT 1770
Db        643 ValTyrTyrThrTyrAlaAsnValTyrIleGlnGluTyrSerSerThrThrIleGluSer 662

Qy       1771 CCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGT 1830
Db        663 GluSerSerThrSerAlaValAlaSerSerThrSerSerThrProSerThrProSerSer 682

Qy       1831 GCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACA 1890
Db        683 ThrLeuSerThrSerThr---ValThrGluProSerSerThrArgSerSerAspSerThr 701

Qy       1891 AGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGA 1950
Db        702 ThrThrSerAlaGlySerThr------ThrThrLeuGlnGluSerThrThrThrSer-Gl 719

Qy       1951 GCCCTGTGTTGAAGATAGTGACACCACAAGT 1981
Db        719 uGluSerThrThrAspSerSerThrThrThr 729
```

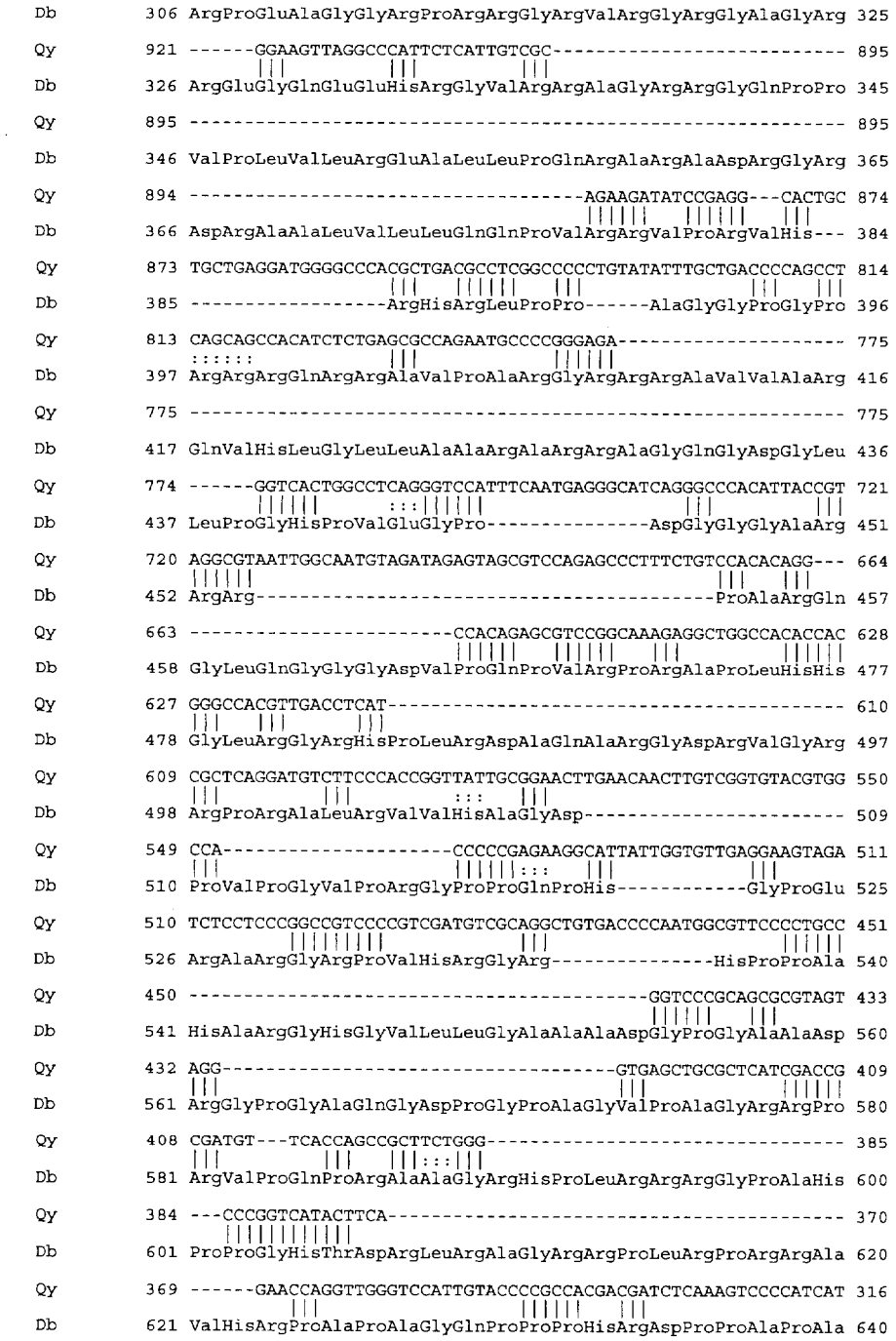
RESULT 15
JQ0405
hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N;Alternate names: ORF 1 protein
C;Species: Micrococcus luteus, Micrococcus lysodeikticus
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C;Accession: JQ0405
R;Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A;Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification of
A;Reference number: S04781; MUID:89364717; PMID:2549377
A;Accession: JQ0405
A;Molecule type: DNA
A;Residues: 1-1106 <SHI>
A;Cross-references: EMBL:X15867
A;Note: this reading frame extends between two stop codons and does not begin with a sta
A;Note: the gene encoding this protein overlaps uvrA gene
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:

Pred. No.:	0.000714	Length:	1106
Score:	199.50	Matches:	202
Percent Similarity:	25.88%	Conservative:	26
Best Local Similarity:	22.93%	Mismatches:	230
Query Match:	4.83%	Indels:	423
DB:	2	Gaps:	49

US-09-914-958B-35 (1-2177) x JQ0405 (1-1106)

```
Qy       1848 CGGGCTTGTCTCGAGGGCACACGAATGGGAACTGGATGCATTCATTGGTGTCCATGCAAT 1789
Db          14 ArgAlaCys------AlaThrValTyrGlyArgGly----------------------- 23

Qy       1788 GGCCATTTTCCTGCT---GGGAGAATCCTTGGCCACACTCCAGTGGGGCTGGGTCCTGAA 1732

Db          24 GlyAspAlaProGlyArgAlaArgLeuSerValGlySerProThrGlyGluThrProVal 43

Qy       1731 GTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGGCCA 1672
Db          44 SerCysPro------------------------------------------------- 46

Qy       1671 CGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCT 1612
Db          47 ---------------------------------CysProProLeuCysTrp--------- 52

Qy       1611 TCCCCAGGCCAAAGTGTGCCACGGGCT-------------------------------- 1585
Db          53 ------------------ProArgAlaGlnGluLeuLeuHisHisArgLeuLeuArgArg 66

Qy       1584 ------CCATCTCACACAGGTAGCCTGAGCCCCCGTCGATGATCCTCAGGT------GGG 1537
Db          67 ArgGlyProArgArgArgProArgLeuArgProArgArgArgAlaLeuGlyArgAlaGly 86

Qy       1536 CCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAA---ACC 1480
Db          87 ProHis---------------ArgArgProGlyArgProArgAlaGlnProGluGlyArg 101

Qy       1479 GGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGAAGCCCTGATTGCCCCGGAAGACGG 1420
Db         102 GlyCysGlnLeuPro--------------------------------AlaGlyArgHis 110

Qy       1419 ACAGCGGCTGAG--------------------------CCATGGACTCTCCATGGGACA 1387
Db         111 GlyArgValHisGlySerValArgLeuGlyGlnValLeuProGlyLeuArgHisAspLeu 130

Qy       1386 AGATGAGGTCCAGCATCCCGTCTCCGTCGA---------------------------- 1357
Db         131 ArgArgGlyProAlaAla---LeuArgArgValAlaLeuLeuLeuArgProHisValPro 149

Qy       1356 ----------------------------AGTCGGTCACCACACCCC------------ 1339
Db         150 GlyProGlyGlyGlnAlaGlyArgGlyLeuHisArgGlyProValProGlyArgValHis 169

Qy       1339 ------------------------------------------------------------ 1339
Db         170 ArgProGluValHisGlnProGlnProAlaLeuHisSerGlyAspHisHisArgAspLeu 189

Qy       1338 ------------------------CTGTGCCCCGGCCCTCAGGCTCCAAGGCGTCGCCGG 1303
Db         190 ArgLeuHisAlaProAlaLeuGlyThrCysArgGlyAlaAlaLeuProAlaValArgArg 209

Qy       1302 --------------------------GATTGAGCTCCTCGATGAGGGGGTCTCCGTGCT 1270
Db         210 AlaGlyGluProAlaAspProAlaAlaAspArgGlyProAlaArgGlyAlaAlaArgAla 229

Qy       1269 ------CTCTACCGATGACGCGGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGT 1216
Db         230 HisProLeuProGlyAlaArgAlaArgGlyPro------ArgProGlnGlyArgVal--- 246

Qy       1215 TGTTGAAGAAGATCTCCAGCTCCTGGT------CATTGTCAAAGTCGGCGGTGATGACCG 1162
Db         247 ---------ArgGlyProValGlnGlyProValHisAlaGlyLeuArgArg--------- 260

Qy       1161 TGCGGACAGGGGAGGGCATGGAGAACTTGGGTGAGGCGATGTCCCGGAAGCGGACCTTCC 1102
Db         261 ------ArgGlyArg----------------------------GlyAspArgProAla 268

Qy       1101 CATGGGTGCTCATTTGCAGATAGAGGCGGTGGGGGCCATTCCAGTTGCCATAGACGATGT 1042
Db         269 LeuGlyProAlaArgAlaGluGluAlaGlyGlnAlaHisHisArg---ArgArgArgGly 287

Qy       1041 CCACTTTGCCATCACGGTTGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGT 982
Db         288 ProProArgHisGluGlyGlyHisProProAlaProHis------ArgLeuGlyGlyAsp 305

Qy        981 CGTCCACACCAGCACTGGCCGCAGCGTCCACAAAGGTGCCATCGCCCCGGTTGTGGAAAA 922
```



Search completed: August 3, 2004, 19:49:42
Job time : 87.5 secs

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:37:19 ; Search time 23 Seconds
(without alignments)
9857.107 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161651_29670/app_query.fasta_1.2375
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914958_@CGN_1_1_24_@runat_03082004_161651_29670 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	272.5	6.7	5703	1	MU5B_HUMAN	Q9hc84 homo sapien
	2	257	6.3	1367	1	AMYH_YEAST	P08640 saccharomyc
	3	253.5	6.2	5179	1	MUC2_HUMAN	Q02817 homo sapien
c	4	223.5	5.4	660	1	YHL1_EBV	P03181 epstein-bar
	5	213.5	5.3	797	1	VGLX_HSVEB	P28968 equine herp
	6	203.5	5.0	3178	1	YS89_CAEEL	Q09624 caenorhabdi
	7	197.5	4.9	907	1	VGP3_EBV	P03200 epstein-bar
	8	197	4.9	1229	1	P121_HUMAN	Q9y2n3 homo sapien
	9	192	4.7	662	1	MUC1_XENLA	Q05049 xenopus lae
c	10	192	4.6	815	1	PYGO_DROME	Q9v9w8 drosophila
	11	191.5	4.7	1199	1	P121_RAT	P52591 rattus norv
	12	191	4.7	676	1	MUC1_MESAU	Q60528 mesocricetu
c	13	190.5	4.6	1185	1	DRPL_HUMAN	P54259 homo sapien
	14	186.5	4.6	1140	1	YM96_YEAST	Q04893 saccharomyc
	15	186	4.6	815	1	PYGO_DROME	Q9v9w8 drosophila
	16	184.5	4.5	1460	1	CA11_CANFA	Q9xsj7 canis famil
	17	184	4.5	1233	1	MU5A_HUMAN	P98088 homo sapien
c	18	182	4.4	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
	19	181.5	4.5	1185	1	DRPL_HUMAN	P54259 homo sapien
	20	180	4.4	1595	1	LTBL_HUMAN	Q14766 homo sapien
	21	179.5	4.4	1453	1	CA11_CHICK	P02457 gallus gall
	22	179.5	4.4	1464	1	CA11_HUMAN	P02452 homo sapien
	23	179	4.4	1161	1	DAN4_YEAST	P47179 saccharomyc
c	24	179	4.3	1509	1	GSR1_HUMAN	Q9nzm4 homo sapien
c	25	178.5	4.3	830	1	SREC_HUMAN	Q14162 homo sapien
	26	178.5	4.4	1418	1	CA12_HUMAN	P02458 homo sapien
	27	178	4.4	784	1	YAV2_XANCV	P14728 xanthomonas
	28	177.5	4.4	623	1	CA44_RABIT	P55787 oryctolagus
	29	177.5	4.4	1336	1	W146_HUMAN	Q9c0j8 homo sapien
	30	176.5	4.4	1446	1	IE18_PRVKA	P33479 pseudorabie
c	31	175.5	4.2	1487	1	ICP4_HSVEK	P17473 equine herp
	32	175	4.3	725	1	AGA1_YEAST	P32323 saccharomyc
	33	174.5	4.3	1461	1	IE18_PRVIF	P11675 pseudorabie
c	34	174.5	4.2	2142	1	BAT2_HUMAN	P48634 homo sapien
	35	172.5	4.3	704	1	FBL1_CHICK	O73775 gallus gall
c	36	172.5	4.2	1487	1	ICP4_HSVEB	P28925 equine herp
	37	171.5	4.2	1048	1	P100_HCMVA	P08318 human cytom
	38	171	4.2	1183	1	DRPL_RAT	P54258 rattus norv
	39	171	4.2	1240	1	YQU3_CAEEL	Q09550 caenorhabdi
	40	171	4.2	1690	1	CA44_HUMAN	P53420 homo sapien
	41	170.5	4.2	892	1	HIC1_MOUSE	Q9r1y5 mus musculu
	42	170.5	4.2	1775	1	CA14_DROME	P08120 drosophila
	43	170.5	4.2	2090	1	N214_HUMAN	P35658 homo sapien
c	44	169.5	4.1	1183	1	DRPL_RAT	P54258 rattus norv
c	45	169.5	4.1	1367	1	AMYH_YEAST	P08640 saccharomyc

ALIGNMENTS

RESULT 1
MU5B_HUMAN
ID MU5B_HUMAN STANDARD; PRT; 5703 AA.
AC Q9HC84; O00447; O00573; O14985; O15494; O95291; O95451; Q14881;
AC Q99552; Q9UE28;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High
DE molecular weight salivary mucin MG1) (Sublingual gland mucin).
GN MUC5B OR MUC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-1594 FROM N.A.
RA Chen Y., Di Y.P., Wu R.;
RT "Molecular cloning of the amino-terminal and 5'-flanking region of the
RT human MUC5B mucin gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 1-1325 FROM N.A.
RX MEDLINE=99009274; PubMed=9790959;
RA Offner G.D., Nunes D.P., Keates A.C., Afdhal N.H., Troxler R.F.;
RT "The amino-terminal sequence of MUC5B contains conserved
RT multifunctional D domains: implications for tissue-specific mucin
RT functions.";
RL Biochem. Biophys. Res. Commun. 251:350-355(1998).
RN [3]
RP SEQUENCE OF 40-1324 FROM N.A.
RX MEDLINE=99023932; PubMed=9804771;
RA Desseyn J.-L., Buisine M.P., Porchet N., Aubert J.-P., Laine A.;
RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic
RT sequences upstream of the large central exon.";
RL J. Biol. Chem. 273:30157-30164(1998).
RN [4]
RP SEQUENCE OF 1326-4895 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97166151; PubMed=9013550;
RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,
RA Laine A.;

```
RT   "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT   various alternate subdomains resulting in a super-repeat. Structural
RT   evidence for a 11p15.5 gene family.";
RL   J. Biol. Chem. 272:3168-3178(1997).
RN   [5]
RP   SEQUENCE OF 4057-4480 FROM N.A.
RC   TISSUE=Salivary gland;
RX   MEDLINE=97292540; PubMed=9147051;
RA   Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,
RA   Hannibal J., Clausen H.;
RT   "Identification of a major human high molecular weight salivary mucin
RT   (MG1) as tracheobronchial mucin MUC5B.";
RL   Glycobiology 7:413-419(1997).
RN   [6]
RP   SEQUENCE OF 4721-5703 FROM N.A.
RC   TISSUE=Gall bladder;
RX   MEDLINE=97293229; PubMed=9164870;
RA   Keates A.C., Nunes D.P., Afdhal N.H., Troxler R.F., Offner G.D.;
RT   "Molecular cloning of a major human gall bladder mucin: complete C-
RT   terminal sequence and genomic organization of MUC5B.";
RL   Biochem. J. 324:295-303(1997).
RN   [7]
RP   SEQUENCE OF 4809-5687 FROM N.A.
RC   TISSUE=Sublingual gland;
RX   MEDLINE=96125355; PubMed=8554565;
RA   Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;
RT   "Molecular cloning of a novel high molecular weight mucin (MG1)
RT   from human sublingual gland.";
RL   Biochem. Biophys. Res. Commun. 217:1112-1119(1995).
RN   [8]
RP   SEQUENCE OF 4859-5703 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97347489; PubMed=9201995;
RA   Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;
RT   "Genomic organization of the 3 region of the human MUC5B mucin.";
RL   J. Biol. Chem. 272:16873-16883(1997).
CC   -!- FUNCTION: Salivary mucin that is thought to contribute to the
CC       lubricating and viscoelastic properties of whole saliva.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also
CC       in submaxillary glands, endocervix, gall bladder, and pancreas.
CC   -!- PTM: Highly glycosylated.
CC   -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC   -!- SIMILARITY: Contains 3 VWFC domains.
CC   -!- SIMILARITY: Contains 4 VWFD domains.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF107890; AAG33673.1; -.
DR   EMBL; AF086604; AAC67545.1; -.
DR   EMBL; AJ004862; CAA06167.1; -.
DR   EMBL; Z72496; CAA96577.1; -.
DR   EMBL; X74955; CAA52910.1; -.
DR   EMBL; U63836; AAB61398.1; -.
DR   EMBL; U78554; AAC51344.1; -.
DR   EMBL; U78552; AAC51344.1; JOINED.
DR   EMBL; U78553; AAC51344.1; JOINED.
DR   EMBL; U78551; AAC51343.1; -.
DR   EMBL; U95031; AAB65151.1; -.
DR   EMBL; Y09788; CAA70926.1; -.
DR   Genew; HGNC:7516; MUC5B.
DR   MIM; 600770; -.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR006208; Cys_knot.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR009041; PMP_inhibitor.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR006552; VC_out.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF01826; TIL; 1.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF00094; vwd; 4.
DR   SMART; SM00214; VWC; 6.
DR   SMART; SM00215; VWC_out; 4.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Glycoprotein; Repeat; Signal; Polymorphism.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   5703       MUCIN 5B.
FT   DOMAIN       77    225       VWFD 1.
FT   DOMAIN      329    386       TIL.
FT   DOMAIN      426    580       VWFD 2.
FT   DOMAIN      858    918       VWFC 1.
FT   DOMAIN      896   1044       VWFD 3.
FT   DOMAIN     1457   1603       THR-RICH.
FT   DOMAIN     1609   4873       THR-RICH.
FT   DOMAIN     5005   5178       VWFD 4.
FT   DOMAIN     5353   5425       VWFC 2.
FT   DOMAIN     5462   5528       VWFC 3.
FT   DOMAIN     5594   5683       CTCK.
FT   DISULFID   5594   5646       BY SIMILARITY.
FT   DISULFID   5622   5660       BY SIMILARITY.
FT   DISULFID   5626   5675       BY SIMILARITY.
FT   DISULFID   5643   5678       BY SIMILARITY.
FT   DISULFID   5645   5682       BY SIMILARITY.
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    516    516       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    806    806       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    930    930       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1277   1277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1293   1293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1557   1557       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1775   1775       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2192   2192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2721   2721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3419   3419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3948   3948       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4745   4745       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4901   4901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4958   4958       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4965   4965       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4987   4987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5037   5037       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5052   5052       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5156   5156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5427   5427       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5467   5467       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5506   5506       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5507   5507       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5543   5543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5553   5553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5604   5604       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5618   5618       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5662   5662       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    5137   5137       T -> S (in dbSNP:2672788).
FT                                /FTId=VAR_014123.
FT   CONFLICT     34     34       G -> E (IN REF. 2).
FT   CONFLICT     95    100       FPGLCN -> LPCLCK (IN REF. 2).
FT   CONFLICT    104    104       S -> C (IN REF. 2).
FT   CONFLICT    142    142       E -> K (IN REF. 1).
FT   CONFLICT    225    225       R -> S (IN REF. 2).
```

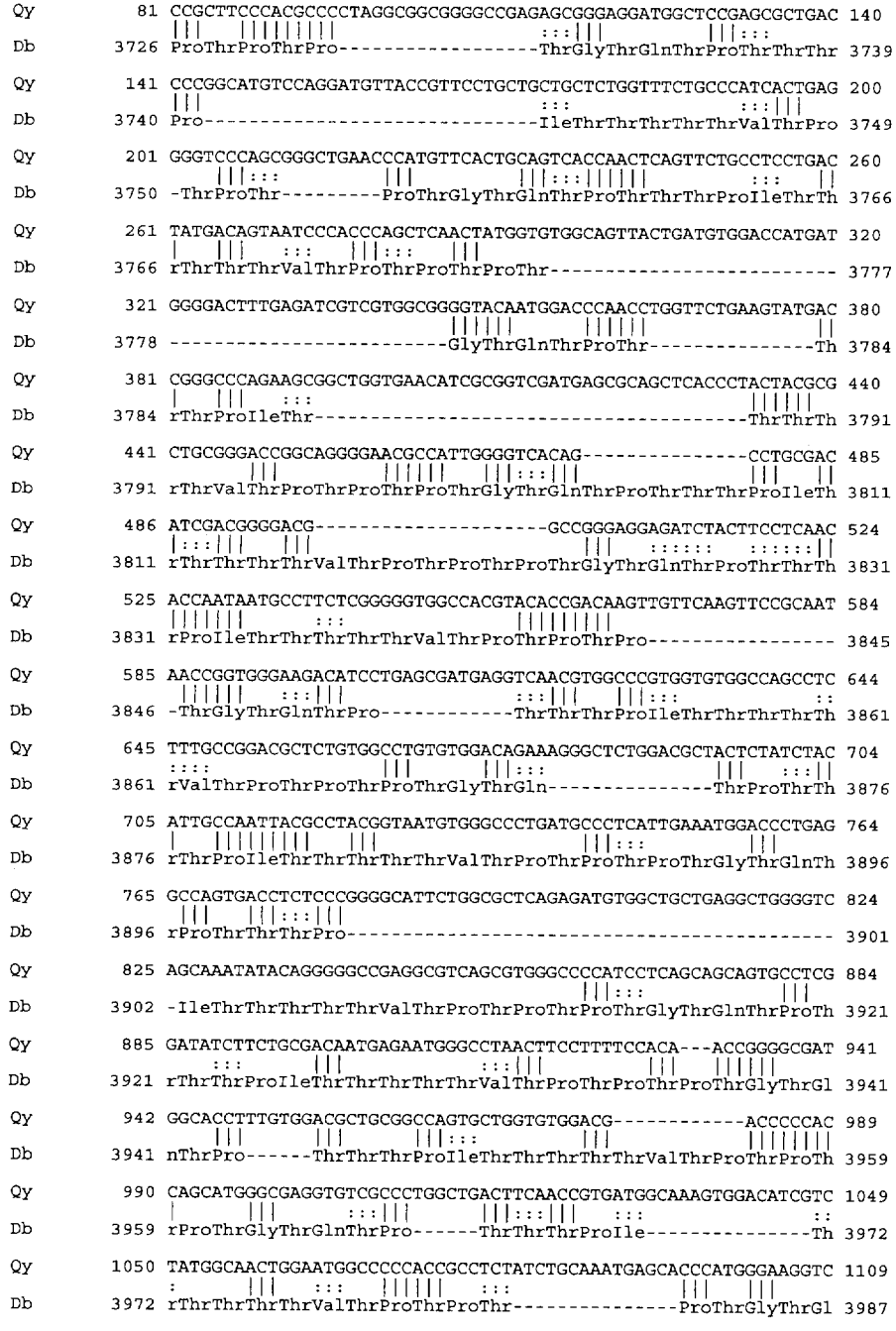
```
FT   CONFLICT    330    331       PL -> T (IN REF. 2).
FT   CONFLICT    337    337       E -> N (IN REF. 2).
FT   CONFLICT    356    356       E -> K (IN REF. 2).
FT   CONFLICT    362    362       G -> R (IN REF. 2).
FT   CONFLICT    369    369       MISSING (IN REF. 2 AND 3).
FT   CONFLICT    374    374       D -> N (IN REF. 2).
FT   CONFLICT    393    394       RT -> TR (IN REF. 2).
FT   CONFLICT    468    469       RK -> GR (IN REF. 2).
FT   CONFLICT    512    512       L -> P (IN REF. 2).
FT   CONFLICT    585    587       GAA -> AH (IN REF. 3).
FT   CONFLICT    601    601       A -> S (IN REF. 3).
FT   CONFLICT    628    629       DP -> RS (IN REF. 2).
FT   CONFLICT    633    633       F -> L (IN REF. 2).
FT   CONFLICT    676    676       A -> P (IN REF. 3).

Alignment Scores:
Pred. No.:                2.14e-07      Length:          5703
Score:                    272.50        Matches:         185
Percent Similarity:       34.41%        Conservative:    91
Best Local Similarity:    23.07%        Mismatches:      269
Query Match:              6.72%         Indels:          257
DB:                       1             Gaps:            39

US-09-914-958B-35 (1-2177) x MU5B_HUMAN (1-5703)

Qy         19 CCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGC 78
Db       4215 ProGlyThrAlaProProProLysValLeuThrSerProAlaThrThrProThrAlaThr 4234

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTG 138
Db       4235 SerSerLysAlaThrSer---SerSerSerProArgThrAlaThrThrLeuProValLeu 4253

Qy        139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA------ 192
Db       4254 ThrSerThrAla------------ThrLysSerThrAlaThrSerValThrProIlePro 4269

Qy        193 ---TCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTC 249
Db       4270 SerSerThrLeuGlyThrThrGlyThrLeuProGluGlnThrThrThrPro--------- 4286

Qy        250 TGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACT------ATGGTGTGGCAGTTA 303
Db       4287 ---ValAlaThrMetSerThrIleHisProSerSerThrProGluThrThrHisThrSer 4305

Qy        304 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACC 363
Db       4306 ThrValLeuThrThrLysAlaThrThrArgAlaThr---SerSerThrSerThrProSer 4324

Qy        364 TGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA 423
Db       4325 --------------SerThrProGlyThrThrTrpIleLeuThrGluLeuThrThrAla 4339

Qy        424 GCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCT--- 480
Db       4340 Ala------ThrThrThrAlaGlyThrGlyProThrAlaThrProSerSerThrProGly 4357

Qy        481 --------------------------GCGACATCGACGGGGACGGCCGGGAGGAGATCT 513
Db       4358 ThrThrTrpIleLeuThrGluLeuThrThrThrAlaThrThrThrAlaSerThrGlySer 4377

Qy        514 ACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGG------------CCACGTACA--- 558
Db       4378 ThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluProSerThrThr 4397

Qy        559 ---------------CCGACA------------AGTTGTTCAAGTTCCGCAATAACCGGT 591
Db       4398 AlaThrValThrValProThrGlySerThrAlaThrAlaSerSerThrGlnAlaThrAla 4417

Qy        592 GGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTG---------------TGG 636
Db       4418 GlyThrProHisValSerThrThrAlaThrThrProThrValThrSerSerLysAlaThr 4437

Qy        637 CCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACT 696
Db       4438 ProSerSerSerProGly------------------ThrAlaThrAlaLeuProAlaLeu 4451

Qy        697 CTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA--------- 747
Db       4452 ArgSerThrAlaThrThrProThrAlaThrSerPheThrAlaIleProSerSerSerLeu 4471

Qy        748 ---TTGAAATGGACC---CTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCA--- 798
Db       4472 GlyThrThrTrpThrArgLeuSerGlnThrThrThrProThrAlaThrMetSerThrAla 4491

Qy        799 --------------------------GAGATGTGGCTGCTGAGGCTGGGGTCAGCAAAT 831
Db       4492 ThrProSerSerThrProGluThrValHisThrSerThrValLeuThrThrThrAlaThr 4511

Qy        832 ATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCA-----------GCAGCAGTG 879
Db       4512 ThrThrGlyAlaThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThr 4531

Qy        880 CCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCG 939
Db       4532 ThrLysValProThrThrThr-----------ThrThrGlyPheThrAlaThrProSer 4547

Qy        940 ATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGG---------------------- 978
Db       4548 SerSerProGlyThrAlaLeuThrProPro---ValTrpIleSerThrThrThrThrPro 4566

Qy        979 ------ACGACCCCCACCAGCATGGGCGAG---GTGTCGCCC------------------ 1011
Db       4567 ThrThrThrThrProThrThrSerGlySerThrValThrProSerSerIleProGlyThr 4586

Qy       1012 ---------------TGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCA 1056
Db       4587 ThrHisThrAlaArgValLeuThrThrThrThrThrThrValAlaThrGlySerMetAla 4606

Qy       1057 ACT-----------------------------------GGAATGGCCCCCACCGCCTCT 1080
Db       4607 ThrProSerSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThr 4626

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
Db       4627 ThrIle--------------------ThrAlaThrGlySerThrThrAsnProSerSer 4639

Qy       1141 ---------------CCATGCCCTCCC---CTGTCCGCACGGTCATCACCGCCGACTTTG 1182
Db       4640 ThrProGlyThrThrProIleProProValLeuThrSerMetAlaThrThrProAlaAla 4659

Qy       1183 ACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCA 1242
Db       4660 ThrSerSerLysAlaThrSerSerSerSerProArgThrAlaThrThrLeuProValLeu 4679

Qy       1243 ACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATC 1302
Db       4680 ThrSerThrAlaThrLysSerThrAlaThrSerPheThrProIleProSerSerThrLeu 4699

Qy       1303 CCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACG 1362
Db       4700 ---TrpThrThrTrpThrValProAlaGlnThrThrThr---------ProMetSerThr 4715

Qy       1363 GAGACGGGATGCTGGACCTCATCTTGTCCC--------------------------ATG 1395
Db       4716 MetSerThrIleHisThrSerSerThrProGluThrThrHisThrSerThrValLeuThr 4735

Qy       1396 GAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACT 1455
Db       4736 ThrThrAlaThrMetThrArgAlaThrAsnSerThrAlaThrProSerSerThrLeu--- 4754

Qy       1456 GGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGC 1515
Db       4755 ----------------------GlyThrThrArgIleLeuThrGluLeuThrThrThr 4766

Qy       1516 TCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGT 1575
```

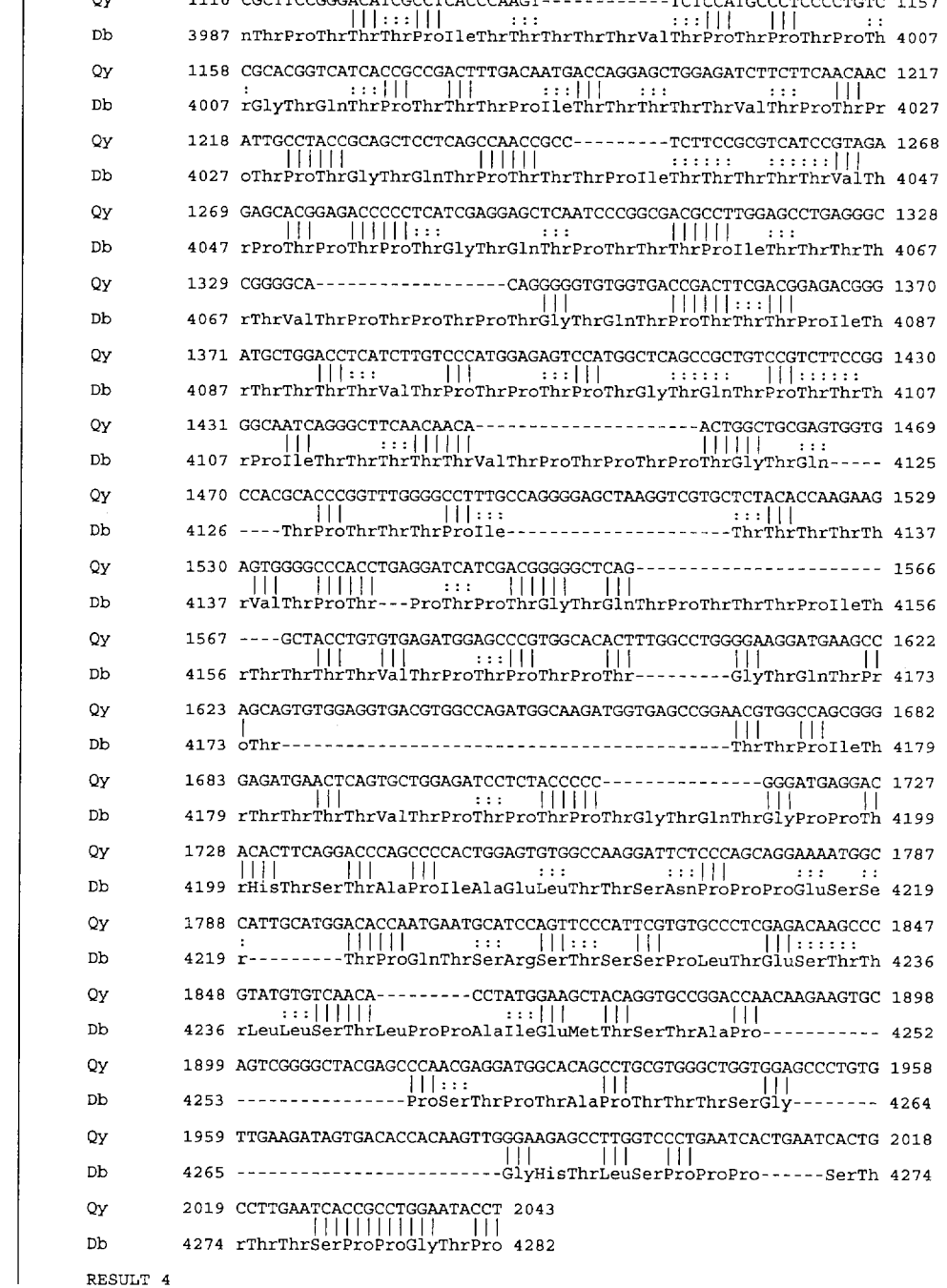
```
Db       4767 AlaThrThrThr--------------------AlaAlaThrGlySerThrAlaThrLeu 4779

Qy       1576 GTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGG 1635

Db       4779 ------------------------------------------------------------ 4779

Qy       1636 TGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAG 1695

Db       4779 ------------------------------------------------------------ 4779

Qy       1696 TGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGT 1755
                    ||||||||||||||||      |||          |||:::|||
Db       4780 ---------SerSerThrProGly---ThrThrTrpIleLeuThrGluPro--------- 4792

Qy       1756 GTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCC 1815
                                      :::|||            |||      ::::::
Db       4793 --------------------SerThrIleAlaThrValMetValProThrGlySerThr 4805

Qy       1816 AGTTCCCATTCGTGTGCCCTCGAGACA------AGCCCGTATGTGTCAACA--------- 1860
              ::::::   |||     |||   |||      :::|||   |||   |||
Db       4806 AlaThrThrSerSerThrLeuGlyThrAlaHisThrProLysValValThrAlaMetAla 4825

Qy       1861 ------CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGC 1914
                    |||                  |||:::   |||   ||||||   |||
Db       4826 ThrMetProThrAlaThrAlaSerThrValProSerSerSerThrValGlyThrThrArg 4845

Qy       1915 CCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACAC 1974

Db       4846 ThrPro-------------------------------------------------------- 4847

Qy       1975 CACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAATCACTGCCTTGAATCACCGCCT 2034
                        |||::::::|||   |||
Db       4848 ---------------AlaValLeuProSerSer---------------------------- 4853

Qy       2035 GGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCACTGAGCAGGATACAAACTTCT 2094
                                  ||||||      |||:::|||   :::   :::|||
Db       4854 ------------------------LeuProThrPheSerValSerThrValSerSerSer 4865

Qy       2095 ATTGTA 2100
              ::::::
Db       4866 ValLeu 4867

RESULT 2
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313266; PubMed=9169870;
RA   Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA   Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA   Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA   Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA   Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL   Nature 387:84-87(1997).
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; S48478; S48478.
DR   GermOnline; 139731; -.
DR   SGD; S0001458; MUC1.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   GO; GO:0030447; P:filamentous growth; IDA.
DR   GO; GO:0007125; P:invasive growth; IMP.
DR   GO; GO:0007124; P:pseudohyphal growth; IMP.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;

Alignment Scores:
Pred. No.:              1.01e-06      Length:          1367
Score:                  257.00        Matches:         182
Percent Similarity:     35.77%        Conservative:    72
Best Local Similarity:  25.63%        Mismatches:      274
Query Match:            6.34%         Indels:          183
DB:                     1             Gaps:            30

US-09-914-958B-35 (1-2177) x AMYH_YEAST (1-1367)

Qy         81 CCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGCTCCGAGCGCTGAC 140
              |||:::|||||||||                |||||         ||||||
Db        313 ProValProThrProSerSerSerThrThrGluSerSer---SerAlaProValProThr 331

Qy        141 CCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAG 200
              |||       |||                                          ||||||
Db        332 ProSerSerSer-----------------------------------------ThrThrGlu 338

Qy        201 GGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGAC 260
                |||      :::   |||:::   ::::::   |||   |||          |||
Db        339 SerSer------SerAlaProValThrSerSerThrThrGluSerSerSerAlaPro-Va 356

Qy        261 TATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGAT 320
                 |||      |||   |||||||||
Db        356 lProThr------ProSerSerSerThr------------------------------- 363

Qy        321 GGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGAC 380
                |||      ||||||         |||         |||         |||   ::
Db        364 ----ThrGluSerSerSerAlaProValThrSerSerThrThrGlu------SerSerSe 380
```

```
Qy       381 CGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCG 440
Db       380 rAlaProValThrSerSer---ThrThrGluSerSerSerAlaProValProThrProSe 399

Qy       441 CTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGC 500
Db       399 rSerSerThr---------ThrGluSerSerSerAlaProValThrSerSerThrThr-- 415

Qy       501 CGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACC 560
Db       416 ----------GluSerSerSerAlaProValThrSerSer------------------Th 426

Qy       561 GACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAAC 620
Db       426 rThrGluSerSerSerAlaProValThrSerSerThrThrGlu---------SerSerSe 443

Qy       621 GTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAG 680
Db       443 rAlaProValThr---SerSerThrThrGluSerSerSerAlaProValProThrProSe 462

Qy       681 GGCTCTGGACGCTACTCTATCTACATTGCCAATTACG---------------------- 717
Db       462 rSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSe 482

Qy       718 ----------CCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACC-------- 759
Db       482 rAlaProValProThr-----------------ProSerSerSerThrThrGluSerSe 496

Qy       760 -CTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCT 818
Db       496 rSerAlaProValThrSerSer---------------------------ThrThrGluSe 507

Qy       819 GGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGT 878
Db       507 rSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProAl 527

Qy       879 GCCTCGGATATCTTCTGCGACAATGAGAATGGGC---------CTAACTTCCTTTTCCAC 929
Db       527 aProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrTh 547

Qy       930 AACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCAC 989
Db       547 rGluSerSerSerAlaProVal---------------------------ProThrProSe 558

Qy       990 CAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTC 1049
Db       558 rSerSerThrThrGluSer---------SerSerThrProValThrSerSerThrThrGl 575

Qy      1050 TATGGCAACTGGAATGGCCCCCACCGCCTCTATC------------TGCAAATGAGCACC 1097
Db       575 uSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaPr 595

Qy      1098 CATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTC 1157
Db       595 oValProThrProSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSe 615

Qy      1158 C---------------GCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGC----- 1197
Db       615 rSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSe 635

Qy      1198 ------------------------------------------------------TG 1199
Db       635 rAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProTh 655

Qy      1200 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 1259
Db       655 rProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerTh 675

Qy      1260 ATCC--------------------------------------------------GT 1265
Db       675 rThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVa 695

Qy      1266 AGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAG 1325
Db       695 lThrSerSerThrThrGluSerSerSerAlaProValPro---ThrProSerSerSerTh 714

Qy      1326 GGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATC 1385
Db       714 rThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSe 734

Qy      1386 TTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTC 1445
Db       734 rAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe 754

Qy      1446 AACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCT 1505
Db       754 rSerThrThr------GluSerSerAlaProValProThrProSerSerSerThrTh 772

Qy      1506 AAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGG------ 1559
Db       772 rGluSerSerSerAlaPro---------ValProThrProSerSerSerThrThrGluSe 789

Qy      1559 ------------------------------------------------------------ 1559
Db       789 rSerSerAlaProValProThrProSerSerSerThrThrGluSerSerValAlaProVa 809

Qy      1560 ------------GGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCT 1606
Db       809 lProThrProSerSerSerSerAsnIleThrSerSerAlaProSerSerThrProPheSe 829

Qy      1607 GGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCG 1666
Db       829 rSerSerThrGluSerSerSerValProValProThrProSerSerSerThrThrGluSe 849

Qy      1667 GAACGTGGCC------------AGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCC 1714
Db       849 rSerSerAlaProValSerSerSerThrThrGluSerSerValAlaProValProThrPr 869

Qy      1715 CCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCA 1774
Db       869 oSerSerSerSerAsnIleThrSerSerAlaProSer-SerIleProPheSerSerThrT 889

Qy      1775 GCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCC 1834
Db       889 hrGluSerPheSerThrGly---ThrThrValThrProSerSerSerLysTyrProGlyS 908

Qy      1835 TCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAA 1894
Db       908 erGlnThrGluThrSerValSerSerThrThrGluThrThr---IleValProThrLysT 927

Qy      1895 GTGCAGTCGGGGCTACGAGCCCAACG 1920
Db       927 hrThrThrSerValThrThrProSer 935
```

```
RESULT 3
MUC2_HUMAN
ID   MUC2_HUMAN     STANDARD;      PRT;  5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mucin 2 precursor (Intestinal mucin 2).
GN   MUC2 OR SMUC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94132002; PubMed=8300571;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT   Identification of the amino terminus and overall sequence similarity
RT   to prepro-von Willebrand factor.";
```

```
RL   J. Biol. Chem. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=93016075; PubMed=1400449;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA   Kim Y.S.;
RT   "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT   both upstream and downstream of its central repetitive region.";
RL   J. Biol. Chem. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE=91358717; PubMed=1885763;
RA   Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA   Petersen G.M., Kim Y.S.;
RT   "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT   and polymorphism.";
RL   J. Clin. Invest. 88:1005-1013(1991).
CC   -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC       other mucus membrane-containing organs. Thought to provide a
CC       protective, lubricating barrier against particles and infectious
CC       agents at mucosal surfaces.
CC   -!- SUBUNIT: Multimeric.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: The number of repeats is highly polymorphic and
CC       varies among different alleles.
CC   -!- SIMILARITY: Contains 1 C-terminal cystine knot-like (CTCK) domain.
CC   -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC   -!- SIMILARITY: Contains 2 VWFC domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   PIR; A49963; A43932.
DR   Genew; HGNC:7512; MUC2.
DR   MIM; 158370; -.
DR   GO; GO:0005803; C:secretory vesicle; TAS.
DR   InterPro; IPR006208; Cys_knot.
DR   InterPro; IPR006207; Cys_knot_C.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF01826; TIL; 1.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF00094; vwd; 4.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   5179       MUCIN 2.
FT   DOMAIN     1401   1747       APPROXIMATE REPEATS.
FT   REPEAT     1401   1416       1.
FT   REPEAT     1417   1432       2.
FT   REPEAT     1433   1448       3.
FT   REPEAT     1449   1464       4.
FT   REPEAT     1465   1471       5.
FT   REPEAT     1472   1478       6.
FT   REPEAT     1479   1494       7A.
FT   REPEAT     1495   1517       7B.
FT   REPEAT     1518   1533       8A.
FT   REPEAT     1534   1556       8B.
FT   REPEAT     1557   1572       9A.
FT   REPEAT     1573   1596       9B.
FT   REPEAT     1597   1612       10A.
FT   REPEAT     1613   1635       10B.
FT   REPEAT     1636   1651       11A.
FT   REPEAT     1652   1675       11B.
FT   REPEAT     1676   1683       12.
FT   REPEAT     1684   1699       13.
FT   REPEAT     1700   1715       14.
FT   REPEAT     1716   1731       15.
FT   REPEAT     1732   1747       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DOMAIN     5075   5160       CTCK.
FT   DISULFID   5075   5122       BY SIMILARITY.
FT   DISULFID   5089   5136       BY SIMILARITY.
FT   DISULFID   5098   5152       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    770    770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1154   1154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1215   1215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1230   1230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1787   1787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1820   1820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4339   4339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4351   4351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4362   4362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4373   4373       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4422   4422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4438   4438       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4502   4502       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4616   4616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4627   4627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4752   4752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4787   4787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4881   4881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4888   4888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4955   4955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4970   4970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5019   5019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5038   5038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5069   5069       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;

Alignment Scores:
Pred. No.:                2.26e-06        Length:          5179
Score:                    253.50          Matches:         160
Percent Similarity:       33.29%          Conservative:    76
Best Local Similarity:    22.57%          Mismatches:      267
Query Match:              6.25%           Indels:          206
DB:                       1               Gaps:            34

US-09-914-958B-35 (1-2177) x MUC2_HUMAN (1-5179)
```





RESULT 4

```
YHL1_EBV
ID   YHL1_EBV       STANDARD;      PRT;   660 AA.
AC   P03181;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical BHLF1 protein.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR   PIR; A03742; QQBE3.
KW   Hypothetical protein; Early protein; Repeat.
FT   DOMAIN      149    648       4 X 125 AA TANDEM REPEATS.
FT   REPEAT      149    273       1.
FT   REPEAT      274    398       2.
FT   REPEAT      399    523       3.
FT   REPEAT      524    648       4.
SQ   SEQUENCE   660 AA;  66244 MW;  86DA1D67A37152A2 CRC64;

Alignment Scores:
Pred. No.:              5.55e-05        Length:         660
Score:                  223.50          Matches:        162
Percent Similarity:     27.30%          Conservative:   28
Best Local Similarity:  23.28%          Mismatches:     218
Query Match:            5.41%           Indels:         288
DB:                     1               Gaps:           34

US-09-914-958B-35 (1-2177) x YHL1_EBV (1-660)

Qy       1965 TCTTCAACACAGGGCTCCACCAGCCCACGCAGGCTGTGCCATCCT-----------CGT 1918
Db        130 SerAlaGlyLeuGlySerArgGlyProArgPro------HisProAlaPheGlnValGln 147

Qy       1917 TGG---GCTCGTAGCCCCGACTGCACTTCTTGTTGGTCCGGCACCTGTAGCTTCCATAGG 1861
Db        148 TrpSerAlaArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArg 167

Qy       1860 TGTTGACACATACGGGCTTGTCTCGAGGGCACACGAATGGGAACTGGATGCATTCATTGG 1801
Db        167 ------------------------------------------------------------ 167

Qy       1800 TGTCCATGCAATGGCCATTTTCCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTG 1741
Db        168 ----------------------------------------GlyHisProProProGlyAla 174

Qy       1740 GGTCCTGAAGTGTGTCCTCATCCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCC 1681
Db        175 GlyGlnArgProSerGlyProThrGlyGlyArg------ProAlaAlaProGlyAlaPro 192

Qy       1680 CGCTGGCCACGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGG 1621
Db        193 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaVal---ProSerGlyAlaThr 211

Qy       1620 CTTCATCCTTCCCCAGGCCAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGCCTGAGC 1561

Db        212 ProHisProGluArgGly---SerGlyProAlaAspProProAlaAlaAlaArgLeuPro 230

Qy       1560 CCCCGTCGATGATCCTCAGGTGGGCCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTC 1501
Db        231 ProGlu----------------------------------ArgGlnGluProArgLeu 238

Qy       1500 CCCTGGCAAAGGCCCCAAACCGGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGAAGC 1441
Db        239 Pro--------------------------GlnAspLeuAlaAla----------Ala 245

Qy       1440 CCTGATTGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGATGA 1381
Db        246 GlnArgCysPro----------AlaGlyProProProThrArgSerGly--------- 258

Qy       1380 GGTCCAGCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCCCTGTGCCCC--------- 1330
Db        259 ------AlaAlaAlaGlnArgThrHisArgArgPro---ProGlyCysProArgSerAla 275

Qy       1329 ---GGCCCTCAGGCTCCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGT 1273
Db        276 ArgAsnProGlyCysProArgThr---------------------------------- 283

Qy       1272 GCTCTCTACGGATGACGCGGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGT 1213
Db        284 ----------------------------TrpArgArgArgSerGlyAlaGln------ 291

Qy       1212 TGAAGAAGATCTCCAGCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAG 1153
Db        292 ---ArgGlyHisProProProGlyAlaGlyGlnArg----------ProSerGlyPro 306

Qy       1152 ------GGGAGG--------------GCATGGAGAACTTGGGTGAGGCGATGTCCCGGA 1114
Db        307 ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 326

Qy       1113 AGCGGA---------CCTTCCCATGGGTGCTCATTTGCAGATAGAGGCGGTGGGGGCCAT 1063
Db        327 GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyProAla 346

Qy       1062 TCC------------------------------------------------------- 1060
Db        347 AspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGlnAsp 366

Qy       1059 ---AGTTGCCATAGACGATGTCCACTTTGCCATCACGGTTGAAGTCAGCCAGGGCGACAC 1003
Db        367 LeuAlaAlaAlaGlnArgCysPro----------------AlaGlyProProProThr 380

Qy       1002 CTCGCCCATGCTGGTGGGGGTCGTCCACACCAGCACTGGCCGCAGCGTCCACAAAGGTGC 943
Db        381 ArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAla 400

Qy        942 CATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATATCCG 883
Db        401 ArgAsnProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisPro 420

Qy        882 AGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTATATTTGCTGA 823
Db        421 Pro------------------------------------------------------- 421

Qy        822 CCCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGAATGCCCCGGGAGAGGTCACTGGCCT 763
Db        422 ----------------------------------------ProGlyAlaGlyGlnArgPro 428

Qy        762 CAGGGTCCATTTCAATGAGGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAATGT 703
Db        429 SerGlyPro---------------------------------------------------- 431

Qy        702 AGATAGAGTAGCGTCCAGAGCCCTTTCTGTCCACACAGGCCACAGAGCGTCCGGCAAAGA 643
Db        432 ---------ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaPro 448

Qy        642 GGCTGGCCACACCACGGGCCACGTTGACCTCATCGCTCAGGATGTCTTCCCACCGGTTAT 583
```

```
Db        449 Gly-----------------------ProGlyGlyGlyAlaAlaValProSerGly--- 459

Qy        582 TGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTATTGGTGT 523
                                          |||   |||   |||
Db        460 --------------------------AlaThrProHisProGluArg------------ 466

Qy        522 TGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGTCGATGTCGCAGGCTGTGACCCCAATGG 463
                    |||   |||      |||           ||||||   |||:::
Db        467 ---------GlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGluArg 483

Qy        462 CGTTCCCCTGCCGGTCCC---------GCAGCGCGTAGTAGGGTGAGCTGCGCTCATCGA 412
                  :::   |||      |||          ||||||      |||
Db        484 GlnGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProPro 503

Qy        411 CCGCGATGTTCACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCAT 352
                |||       |||   ||||||:::          |||         ||||||
Db        504 ProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGly--------- 520

Qy        351 TGTACCCCGCCACGACGATCTCAAAGTCCCCATCATGGTCCACATCAGTAACTGCCACAC 292
                |||       ||||||   :::::::::|||
Db        521 Cys------ProArgSerAlaArgAsnPro---------------------------- 528

Qy        291 CATAGTTGAGCTGGGTGGGATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTG 232

Db        528 ------------------------------------------------------------ 528

Qy        231 CAGTGAACATGGGTTCAGCCCGCTGGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCA 172
                                  |||   |||:::   |||                |||
Db        529 ------------------------GlyCysProArgThrTrpArgArgArgSerGlyAla 540

Qy        171 GCAGGAACGGTAACATCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCT 112
                                          ||||||      ||| ||||||
Db        541 --------------------------------GlnArgGlyHisPro-ProProGlyAl 549

Qy        111 CGGCCCCGCCGCCTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCGCCTCTGCCGCCGGCGC 52
                |||       |||   |||         ||||||      |||      |||   ||||
Db        549 aGlyGlnArg-ProSerGlyProThrGlyGlyArg------ProAlaAlaProGlyAlaP 567

Qy         51 CG---CCGCCTGCTTGCTCCCAGCCCCGGTCCCGGGCTGGCC 13
                ||      |||||| ||||||   ||||||   |||   |||
Db        567 roGlyThrProAla-AlaProGlyProGlyGlyGlyAlaAla 580

RESULT 5
VGLX_HSVEB
ID   VGLX_HSVEB     STANDARD;      PRT;   797 AA.
AC   P28968;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   Glycoprotein X precursor.
GN   71.
OS   Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=31520;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295566; PubMed=1318606;
RA   Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86664; AAB02506.1; -.
DR   PIR; H36802; VGBEX1.
DR   InterPro; IPR007110; Ig-like.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    797       GLYCOPROTEIN X.
FT   DOMAIN       23    465       SER/THR-RICH.
FT   TRANSMEM    766    790       POTENTIAL.
FT   CARBOHYD    590    590       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   797 AA;  80342 MW;  50C9ED9211F5E5B2 CRC64;

Alignment Scores:
Pred. No.:                0.000205      Length:          797
Score:                    213.50        Matches:         113
Percent Similarity:       35.70%        Conservative:    68
Best Local Similarity:    22.29%        Mismatches:      197
Query Match:              5.26%         Indels:          129
DB:                       1             Gaps:            15

US-09-914-958B-35 (1-2177) x VGLX_HSVEB (1-797)

Qy        403 ACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACG 462
                |||:::   :::   :::              ::::::      |||:::|||:::   |||
Db         23 ThrThrThrThrGluThrThrThrSerSerSerSerThrSerGlySerGlyGlnSerThr 42

Qy        463 CCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCA 522
                  |||:::      ::::::|||:::   |||            :::|||||||||
Db         43 SerSerGlyThrThrAsnSerSerSerSerProThrThrSerProProThrThrSerSer 62

Qy        523 ACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCA 582
                :::|||      |||:::                 :::|||:::|||   |||:::   :::
Db         63 SerPro------ProThrSerThrHisThrSerSerProSerSerThrSerThrGlnSer 80

Qy        583 ATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCC 642
                  :::         ||||||   :::      ||||||                  |||
Db         81 SerSerThrAlaAlaThrSerSerSerAlaProSerThr------------------Ala 94

Qy        643 TCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCT 702
                |||         ::::::   |||      |||                     |||
Db         95 SerSerThrThrSerIle---ProThrSerThrSerThrGluThrThrThrThrPro 113

Qy        703 ACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTG 762
                |||         |||                          |||:::        |||
Db        114 ThrAlaSerThrThrThrProThrThrThrThrAlaAlaProThrThrAlaAlaThrThr 133

Qy        763 AGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGG 822
                  ||||||:::
Db        134 ThrAlaValThrThr----------------------------------------AlaAla 140

Qy        823 TCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCT 882
                |||   :::   :::          |||:::|||   |||   |||:::
Db        141 SerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSerThrProThrThrThr 160

Qy        883 CGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATG 942
                    :::|||   |||                |||:::      ||||||   :::
Db        161 ThrProThrSerThrThrThrThrThrAlaThrThrThrValProThrThrAlaSerThr 180

Qy        943 GCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAG 1002
                      |||                         |||   |||:::
Db        181 ThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 200

Qy       1003 GTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGA 1062
                  :::         |||:::                |||::::::   ||||||
Db        201 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 220

Qy       1063 ATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACA 1122
                |||   |||   |||                                  |||   |||
Db        221 ThrAlaAlaThrThrSer---------------------------------SerAlaThr 229

Qy       1123 TCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTG 1182
                :::         ::::::              :::|||   ::::::      |||
Db        230 ThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThr 249

Qy       1183 ACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCA 1242
                |||                  :::::::::||||||      |||           |||
Db        250 Thr-----------------AlaAlaThrThrThrAlaAlaThrThrThrGlySerPro 263
```

```
Qy       1243 ACCGCCTCTTCCGCGTCATCCGTA---GAGAGCACGGAGACCCCCTCATCGAGGAGCTCA 1299
                 |||:::   |||   |||:::         ||||||   |||||||||:::   ::::::
Db        264 ThrSerGlySerThrSerThrThrGlyAlaSerThrSerThrProSerAlaSerThrAla 283

Qy       1300 ATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCG 1359
                 |||||||||   |||                                     ||||||
Db        284 ThrSerAlaThrProThrSer---------------------------------ThrSer 292

Qy       1360 ACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGT 1419
                 |||                |||||                          :::
Db        293 ThrSerAlaAlaAlaThrThrSer-------------------------ThrProThr 303

Qy       1420 CCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCC 1479
                 |||:::|||   |||      |||   :::||||||                  ||||||
Db        304 ProThrSerAlaAlaThrSerAlaGluSerThrThr--------------GluAlaPro 318

Qy       1480 GGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCC 1539
                                              ||||||               |||
Db        319 ---------------------------------ThrSerThr--------------Pro 322

Qy       1540 ACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACT 1599
                 |||      :::::::|||   ::::::||||||            ||||||      |||
Db        323 ThrThrAspThrThrThrProSerGluAlaThrThrAlaThrThrSerProGluSerThr 342

Qy       1600 TTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGG 1659

Db        343 ThrValSerAla---------------------------------------------- 346

Qy       1660 TGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGG 1719
                                                          |||:::|||
Db        347 ---------------------------------------------SerThrThrSerAla 351

Qy       1720 ATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGG 1779
                 |||   |||   |||:::   |||   |||
Db        352 ThrThrThrAlaPheThrThrGluSerHisThrSer------------------------ 363

Qy       1780 AAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAG 1839
                              |||   :::::::::   |||   |||   |||
Db        364 --------------------ProAspSerSerThrGlySerThrSerThrAlaGluPro 376

Qy       1840 ACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCA 1899
                 :::|||   :::      ||||||      ||||||   :::      |||
Db        377 SerSerThrPheThrLeuThrProSerThrAlaThrProSer------ThrAspGlnPhe 394

Qy       1900 GTCGGGGCTACGAGCCCAACG 1920
                 |||:::::::::   |||
Db        395 ThrGlySerSerAlaSerThr 401
```

```
RESULT 6
YS89_CAEEL
ID   YS89_CAEEL     STANDARD;      PRT;  3178 AA.
AC   Q09624; Q09625; Q969D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   -!- SIMILARITY: Contains 1 PLAT domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; JOINED.
DR   EMBL; Z48582; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
DR   WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR003915; PKD_2.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR008976; PLAT_LH2.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   PRINTS; PR01433; POLYCYSTIN2.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
DR   PROSITE; PS50095; PLAT; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     13     30       POTENTIAL.
FT   TRANSMEM     51     73       POTENTIAL.
FT   TRANSMEM   2139   2161       POTENTIAL.
FT   TRANSMEM   2348   2367       POTENTIAL.
FT   TRANSMEM   2390   2412       POTENTIAL.
FT   TRANSMEM   2451   2468       POTENTIAL.
FT   TRANSMEM   2483   2505       POTENTIAL.
FT   TRANSMEM   2567   2589       POTENTIAL.
FT   TRANSMEM   2836   2858       POTENTIAL.
FT   TRANSMEM   2939   2961       POTENTIAL.
FT   TRANSMEM   2976   2998       POTENTIAL.
FT   TRANSMEM   3038   3060       POTENTIAL.
FT   DOMAIN      266   1196       SER/THR-RICH.
FT   DOMAIN     1105   1241       GLY/SER-RICH.
FT   DOMAIN     2071   2120       GPS.
FT   DOMAIN     2182   2305       PLAT.
SQ   SEQUENCE   3178 AA;  344726 MW;  F8239436D03666CD CRC64;
```

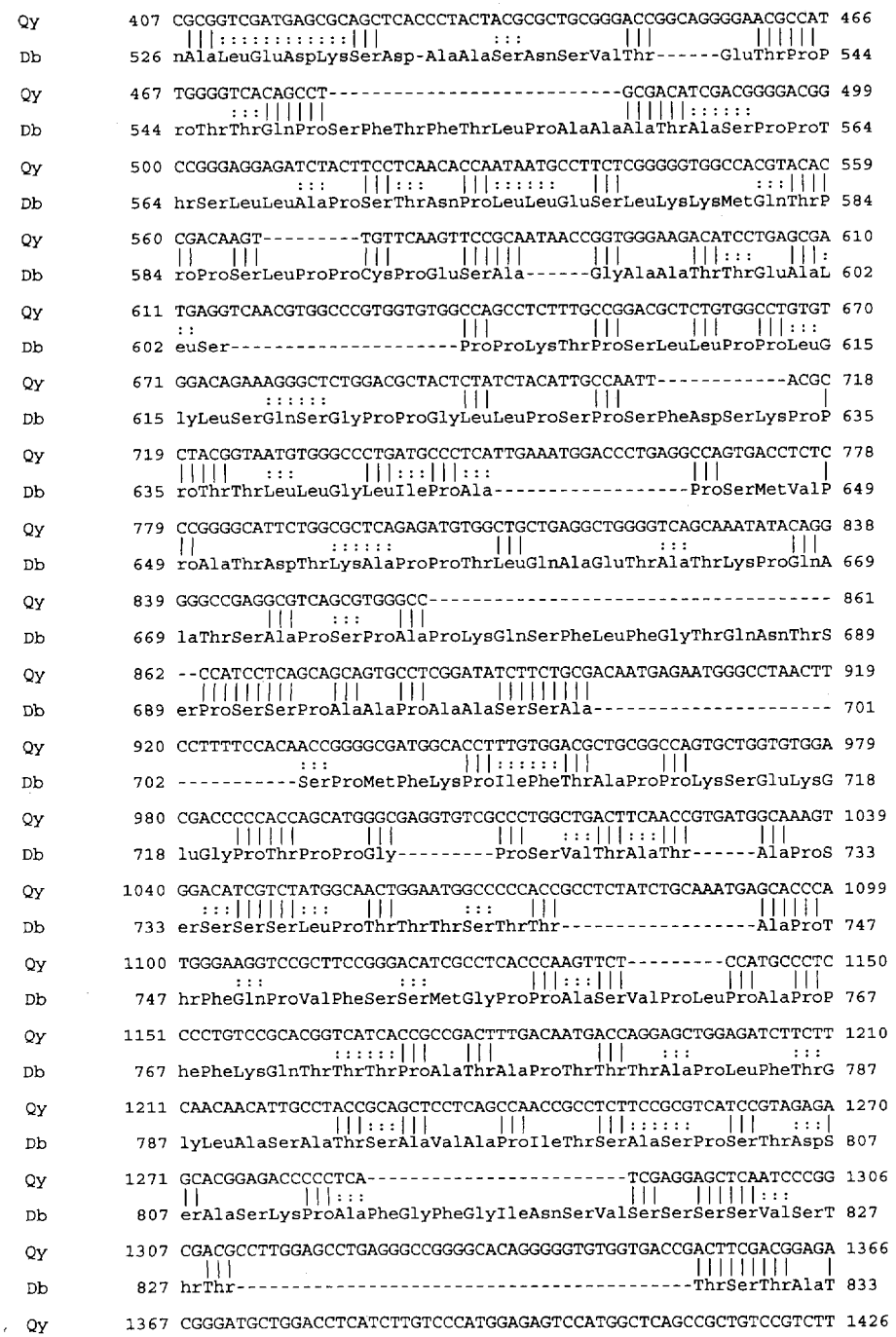
```
Alignment Scores:
Pred. No.:                0.00105        Length:          3178
Score:                    203.50         Matches:         151
Percent Similarity:       36.97%         Conservative:    83
Best Local Similarity:    23.85%         Mismatches:      210
Query Match:              5.02%          Indels:          189
DB:                       1              Gaps:            28

US-09-914-958B-35 (1-2177) x YS89_CAEEL (1-3178)

Qy        190 CCATCACTGAGG---------GGTCCCAGCGGGCTGAACCCATGT------TCACTGCAG 234
                  |||:::||||||               :::|||   |||   |||      :::::::::
Db        253 ProThrLeuArgArgMetLysArgAspAlaGlyAspAsnThrCysAspTyrThrIleGlu 272

Qy        235 TCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTG 294
                 |||   :::                |||   |||       |||   ::::::|||:::
Db        273 SerThrSer---------------ThrSerThrThrThrProThrThrThrThrValThr 287

Qy        295 TGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATG 354
                                           |||:::   |||:::         |||
Db        288 --------------------------SerThrValThrSerThr---------ThrThr 295

Qy        355 GACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCG 414
                 ||||||         ::::::|||                        |||:::   |||
Db        296 ValProThrSerThrSerThrValThr--------------------ThrAlaMetSer 308
```

```
Qy       415 ATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCA 474
Db       309 ThrSerThrSerThrProSerThr-----------------SerThrThrIleGluSer 322

Qy       475 CAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATG 534
Db       323 ThrSerThrThrPheThrSerThrAlaSer---ThrSerThrSerSerThr--------- 338

Qy       535 CCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGG 594
Db       339 --------------------SerThrThrGlnGlnSerSerSerThrIleThrSerSer 351

Qy       595 AAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC 654
Db       352 ProSerSer------------------------------------------------- 354

Qy       655 GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACA---TTGCCA 711
Db       355 -----------------------------------ThrThrLeuSerThrSerIlePro 362

Qy       712 ATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTG 771
Db       363 ThrThrThrThr-----------------ProGluIleThrSerThrLeuSerSerLeu 376

Qy       772 ACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAAT 831
Db       376 ------------------------------------------------------------ 376

Qy       832 ATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCT 891
Db       377 ----------------------------ProAspAsnAlaIleCysSerTyrLeuAsp 386

Qy       892 TCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTG 951
Db       387 GluThrThrThrSer---------ThrThrPheThrThrThr------------------ 397

Qy       952 TGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCC 1011
Db       398 ------------------------MetLeuThrSerThrThrGluGluProSerThr 409

Qy      1012 TGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCC 1071
Db       410 SerThrThrThrThrGluValThrSer---ThrSerSerThrValThrThrThrGluPro 428

Qy      1072 ACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCAC 1131
Db       429 ThrThrThrLeuThrThr--------------SerThrAlaSerThrSerThrThrGlu 443

Qy      1132 CCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACC 1191
Db       444 ProSerThrSerThrValThrThrSerProSerThrSerProValThrSerThrValThr 463

Qy      1192 AGGAGCTGGAGATCTTCTTCAACA---ACATTGCCTACCGCAGCTCCTCAGCCAACCGCC 1248
Db       464 SerSerSerSerSerSerThrThrValThrThrProThrSerThrGluSerThrSerThr 483

Qy      1249 TCTTCCGCGTCATCCGTA---GAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCG 1305
Db       484 SerProSerSerThrValThrThrSerThrThrAlaProSerThrSerThrThrGlyPro 503

Qy      1306 GCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAG 1365
Db       504 SerSerSerSerSerThr--------------------------ProSerSerThrAla 514

Qy      1366 ACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCT 1425
Db       515 SerSerSerValSerSerThrAlaSerSerThrGlnSerSerThrSerThrGlnGlnSer 534

Qy      1426 TCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTG 1485
Db       535 SerThrThrThrLysSerGluThrThrThr-SerSerAspGly---ThrAsnProAspPh 553

Qy      1486 GGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT------------ 1532

Db       553 e-TyrPheValGluLysAlaThrThrThrPheTyrAspSerThrSerValAsnLeuThrL 573

Qy      1533 --GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCG 1590
Db       573 euAsnSerGlyLeuGlyIleIle------------GlyTyrGlnThrSerIleGluCysT 589

Qy      1591 TGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATG 1650
Db       589 hrSer------------------ProThrSerSerAsnTyrValSerThrThrLysAspG 603

Qy      1651 GCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCT 1710
Db       603 lyAlaCysPheThrLysSerValSerMetProArgLeuGlyGlyThr-TyrProAlaSer 622

Qy      1711 ACC---------CCCGGGATGAGGACACACTTCAGGACC---------------------- 1740
Db       623 ThrPheValGlyProGlyAsnTyrThrPheArgAlaThrMetThrThrAspAspLysLys 642

Qy      1741 ------------------------------CAGCCCCACTGGAGTGTGGCCAAGGATTCT 1770
Db       643 ValTyrTyrThrThrTyrAlaAsnValTyrIleGlnGluTyrSerSerThrThrIleGluSer 662

Qy      1771 CCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGT 1830
Db       663 GluSerSerThrSerAlaValAlaSerSerThrSerSerThrProSerThrProSerSer 682

Qy      1831 GCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACA 1890
Db       683 ThrLeuSerThrSerThr---ValThrGluProSerSerThrArgSerSerAspSerThr 701

Qy      1891 AGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGA 1950
Db       702 ThrThrSerAlaGlySerThr------ThrThrLeuGlnGluSerThrThrThrSer-Gl 719

Qy      1951 GCCCTGTGTTGAAGATAGTGACACCACAAGT 1981
Db       719 uGluSerThrThrAspSerSerThrThrThr 729
```

```
RESULT 7
VGP3_EBV
ID   VGP3_EBV       STANDARD;      PRT;   907 AA.
AC   P03200; P03201;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Envelope glycoprotein GP340/GP220 (Membrane antigen) (MA).
GN   BLLF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS GP220 AND GP340).
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=GP340;
CC         IsoId=P03200-1; Sequence=Displayed;
CC       Name=GP220;
CC         IsoId=P03200-2; Sequence=VSP_002070;
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; V01555; CAA24854.1; -.
DR   PIR; A43042; QQBE21.
DR   PIR; B43042; QQBE22.
DR   InterPro; IPR007796; Herpes_BLLF1.
DR   Pfam; PF05109; Herpes_BLLF1; 1.
KW   Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    610    610       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    701    701       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    735    735       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    746    746       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    755    755       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    815    815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    858    858       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    888    888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    502    698       Missing (in isoform GP220).
FT                                /FTId=VSP_002070.
SQ   SEQUENCE   907 AA;  94431 MW;  0750141CBCAC52C9 CRC64;

Alignment Scores:
Pred. No.:                0.00159        Length:          907
Score:                    197.50         Matches:         126
Percent Similarity:       36.47%         Conservative:    56
Best Local Similarity:    25.25%         Mismatches:      216
Query Match:              4.87%          Indels:          102
DB:                       1              Gaps:            23

US-09-914-958B-35 (1-2177) x VGP3_EBV (1-907)

Qy        600 ATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC----- 654
Db        304 ValPheSerAspGluIleProAlaSerGlnAspMet-ProThrAsnThrThrAspIleTh 323

Qy        655 ----------------GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCT 698
Db        323 rTyrValGlyAspAsnAlaThrTyrSerValProMetValThrSerGluAspAla----- 341

Qy        699 ATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA----------- 747

Db        342 -AsnSerProAsnValThrValThrAlaPheTrpAla---TrpProAsnAsnThrGluTh 360

Qy        748 ----------TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTC 797
Db        360 rAspPheLysCysLysTrpThrLeuThrSerGlyThrProSerGlyCys---------- 376

Qy        798 AGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCA----- 852
Db        377 ----------------------------GluAsnIleSerGlyAlaPheAlaSerAsnAr 387

Qy        853 ----------------------------GCGTGGGCCCCATCCTCAGCAGCAGTGCCTCG 884
Db        387 gThrPheAspIleThrValSerGlyLeuGlyThrAlaProLysThrLeuIleIleThrAr 407

Qy        885 GATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGC 944
Db        407 gThrAlaThrAsn----------------------------AlaThrThrThrThrHisLy 418

Qy        945 ACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACC------AGCATGGG 998
Db        418 sValIlePheSerLysAlaProGluSerThrThrThrSerProThrLeuAsnThrThrGl 438

Qy        999 CGAGGTGTCGCCCTGGCTGACT-----------------TCAACCGTGATGGCAAAGTG 1040
Db        438 yPheAlaAspProAsnThrThrThrGlyLeuProSerSerThrHisValProThrAsnLe 458

Qy       1041 GACATCGTCTATGGCAACTGGA---ATGGCCCCCACCGCCTCTATCTGCAAATGAGCACC 1097
Db        458 uThrAlaProAlaSerThrGlyProThrValSerThrAlaAspValThrSerProThrPr 478

Qy       1098 CATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCT---CCCCT 1154
Db        478 oAlaGly---ThrThrSerGlyAlaSerProValThrProSerProSerProTrpAspAs 497

Qy       1155 GTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAAC 1214
Db        497 nGlyThrGluSerLysAlaProAspMetThr----------------------SerSe 509

Qy       1215 AACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCAC 1274
Db        509 rThrSerProValThrThrProThrProAsnAlaThrSerProThrProAlaValThrTh 529

Qy       1275 GGAGACCCCCTCATCGAGGAGCTCAATCCCGGCG------------ACGCCTTGGAGCCT 1322
Db        529 rProThrProAsnAlaThrSerProThrProAlaValThrThrProThrPro---AsnAl 548

Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382
Db        548 aThrSerProThrLeuGlyLysThrSerProThrSerAlaValThr---------ThrPr 565

Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGC 1442
Db        565 oThrProAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaValThrTh 585

Qy       1443 TTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTG---CCAGG 1499
Db        585 rProThrProAsnAlaThr-----------SerProThrLeuGlyLysThrSerProTh 601

Qy       1500 GGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGG 1559
Db        601 rSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGly---GluThrSe 620

Qy       1560 GGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAA 1619
Db        620 rProGlnAlaAsnAlaThrAsn-----------HisThrLeuGlyGlyThrSerProTh 636

Qy       1620 GCCAGCAGTGTG------------------------GAGGTGACGTGGCCAGATGGCAA 1654
Db        636 rProValValThrSerGlnProLysAsnAlaThrSerAlaValThrThrGlyGlnHisAs 656

Qy       1655 GATGGTGAGCCCGAACGTGGCCAGCGGGGAGATG------AACTCAGTGCTGGAGATCCT 1708
```

```
Db        656 nIleThrSerSerSerThrSerSerMetSerLeuArgProSerSerAsnProGluThrLe 676

Qy       1709 CTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATT 1768
              |   |||         |||                     ||||||
Db        676 uSerProSerThrSerAspAsnSerThrSerHisMetProLeuLeuThrSerAlaHisPr 696

Qy       1769 CTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGT 1828
               :::      ||||||              ||||||   :::   |||:::   |||
Db        696 oThrGlyGlyGluAsn-IleThrGlnValThrProAlaSerIleSerThrHisHisValS 716

Qy       1829 GTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA---- 1884
                          :::|||          |||      :::|||:::|||   |||
Db        716 erThrSerSerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProGlyAsnS 736

Qy       1885 --CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCT 1935
                 |||   :::      |||      :::   |||:::          |||
Db        736 erSerThrSerThrLysProGlyGluValAsnValThrLysGlyThrProPro 753

RESULT 8
P121_HUMAN
ID   P121_HUMAN     STANDARD;      PRT;  1229 AA.
AC   Q9Y2N3; O75115; Q9Y4S7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE   of 121 kDa) (P145).
GN   POM121 OR NUP121 OR KIAA0618.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cordes M., Bauer C., Holmes A.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 243-1229 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE OF 1130-1229 FROM N.A.
RC   TISSUE=Uterus;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential component of the nuclear pore complex. The
CC       repeat-containing domain may be involved in anchoring components
CC       of the pore complex to the pore membrane (By similarity).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC       MEMBRANE (BY SIMILARITY).
CC   -!- DOMAIN: Contains F-X-F-G repeats.
CC   -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC       SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC       NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC006014; AAD28064.1; -.
DR   EMBL; AB014518; BAA31593.1; -.
DR   EMBL; AL080109; CAB45713.1; -.
DR   PIR; T12551; T12551.
DR   Genew; HGNC:19702; POM121.
KW   Nuclear protein; Transport; Transmembrane; Repeat.
FT   DOMAIN        1     40       CISTERNAL SIDE (POTENTIAL).
FT   TRANSMEM     41     61       POTENTIAL.
FT   DOMAIN       62   1229       PORE SIDE (POTENTIAL).
FT   DOMAIN        4     10       POLY-ALA.
FT   DOMAIN       51     56       POLY-ALA.
FT   DOMAIN      294    299       POLY-LYS.
FT   DOMAIN      441    444       POLY-SER.
FT   DOMAIN      499    502       POLY-PRO.
FT   DOMAIN      733    736       POLY-SER.
FT   DOMAIN      819    826       POLY-SER.
FT   DOMAIN      869    875       POLY-THR.
FT   DOMAIN     1061   1067       POLY-SER.
FT   CONFLICT    297    297       K -> E (IN REF. 2).
FT   CONFLICT    336    336       S -> N (IN REF. 2).
FT   CONFLICT    379    379       T -> A (IN REF. 2).
FT   CONFLICT    451    451       K -> R (IN REF. 2).
FT   CONFLICT    545    545       T -> I (IN REF. 2).
FT   CONFLICT    559    559       T -> P (IN REF. 2).
FT   CONFLICT    702    702       S -> P (IN REF. 2).
FT   CONFLICT    881    881       P -> H (IN REF. 2).
FT   CONFLICT    965    965       T -> A (IN REF. 2).
FT   CONFLICT    982    984       PAT -> A (IN REF. 2).
FT   CONFLICT    991    996       ASTIKI -> PSMIKV (IN REF. 2).
FT   CONFLICT   1000   1000       H -> Y (IN REF. 2).
FT   CONFLICT   1006   1008       QPT -> HPI (IN REF. 2).
FT   CONFLICT   1044   1044       MISSING (IN REF. 2).
FT   CONFLICT   1102   1102       A -> T (IN REF. 2).
FT   CONFLICT   1124   1124       T -> A (IN REF. 2).
FT   CONFLICT   1146   1146       S -> G (IN REF. 1).
FT   CONFLICT   1165   1165       L -> Q (IN REF. 1).
FT   CONFLICT   1195   1195       A -> G (IN REF. 1).
FT   CONFLICT   1199   1229       SAAPSFSIGAGSKTPGARQRLQARRQHTRKK -> NTFAHQ
FT                                QEHSPRKGPNNLSKRKLLPAVRAQGPPRRGQASSFPTRKE
FT                                (IN REF. 2).
FT   CONFLICT   1202   1202       P -> L (IN REF. 3).
SQ   SEQUENCE   1229 AA;  125087 MW;  515655D12858998B CRC64;

Alignment Scores:
Pred. No.:                0.00184        Length:          1229
Score:                    197.00         Matches:         164
Percent Similarity:       33.10%         Conservative:    76
Best Local Similarity:    22.62%         Mismatches:      245
Query Match:              4.86%          Indels:          240
DB:                       1              Gaps:            33

US-09-914-958B-35 (1-2177) x P121_HUMAN (1-1229)

Qy         47 CGGCGGCGCCGGCGGCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGG 106
              |||:::::::::            |||      |||   |||         |||:::
Db        464 ArgLysLysGlnAsnSerAsnSerGlnSerThrProGlySerSerGlyGlnArgLys--- 482

Qy        107 GGCCGAGAGCGGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTT 166
                                                  |||   |||||| :::||||||
Db        483 ----------------------------------ArgLysValGln-LeuLeuPro-- 489

Qy        167 CCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTT 226
                                                 |||:::|||   |||   :::
Db        490 ----------------------------------SerArgArgGlyGluGlnLeu-- 496

Qy        227 CACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAA 286
                                      |||||||||           |||   ||||||
Db        497 ------------------ThrLeuProPro--------------ProProGlnLeuGl 505

Qy        287 CTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGG 346
              |||   :::          |||:::|||   :::
Db        505 yTyrSerIleThrAlaGluAspLeuAspLeuGlu------------------------- 516

Qy        347 GTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACAT 406
                                             :::|||      :::      |||
Db        517 ------------------------------LysLysAlaSerLeuGlnTrpPheAsnGl 526
```



RESULT 9
MUC1_XENLA
ID MUC1_XENLA STANDARD; PRT; 662 AA.
AC Q05049;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integumentary mucin C.1 (FIM-C.1) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6 AND 7).
RC TISSUE=Skin;
RX MEDLINE=93077556; PubMed=1447205;
RA Hauser F., Hoffmann W.;
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT polymorphism.";
RL J. Biol. Chem. 267:24620-24624(1992).

```
CC   -!- FUNCTION: Could be involved in defense against microbial
CC       infections. Protects the epithelia from external environment.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=7;
CC         Comment=Additional isoforms seem to exist. Experimental
CC         confirmation may be lacking for some isoforms;
CC       Name=1;
CC         IsoId=Q05049-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q05049-2; Sequence=VSP_004650;
CC       Name=3;
CC         IsoId=Q05049-3; Sequence=VSP_004651;
CC       Name=4;
CC         IsoId=Q05049-4; Sequence=VSP_004647, VSP_004648;
CC       Name=5;
CC         IsoId=Q05049-5; Sequence=VSP_004646, VSP_004649, VSP_004650;
CC       Name=6;
CC         IsoId=Q05049-6; Sequence=VSP_004646, VSP_004648;
CC       Name=7;
CC         IsoId=Q05049-7; Sequence=VSP_004647;
CC   -!- TISSUE SPECIFICITY: Skin.
CC   -!- PTM: Extensively O-glycosylated.
CC   -!- SIMILARITY: Contains 6 P-type (trefoil) domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L02115; AAA74725.1; -.
DR   PIR; A45155; A45155.
DR   HSSP; P01359; 2PSP.
DR   InterPro; IPR000519; P_trefoil.
DR   Pfam; PF00088; trefoil; 6.
DR   PRINTS; PR00680; PTREFOIL.
DR   SMART; SM00018; PD; 6.
DR   PROSITE; PS00025; P_TREFOIL; 6.
KW   Repeat; Glycoprotein; Alternative splicing.
FT   NON_TER       1      1
FT   DOMAIN       81    144       8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT                                ALA/THR-RICH.
FT   REPEAT       81     88       1-1.
FT   REPEAT       89     96       1-2.
FT   REPEAT       97    104       1-3.
FT   REPEAT      105    112       1-4.
FT   REPEAT      113    120       1-5.
FT   REPEAT      121    128       1-6.
FT   REPEAT      129    136       1-7.
FT   REPEAT      137    144       1-8.
FT   DOMAIN      161    202       P-TYPE 1.
FT   DOMAIN      218    301       8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT   REPEAT      218    224       2-1.
FT   REPEAT      225    239       2-2.
FT   REPEAT      240    249       2-3.
FT   REPEAT      250    259       2-4.
FT   REPEAT      260    275       2-5.
FT   REPEAT      276    287       2-6.
FT   REPEAT      288    294       2-7.
FT   REPEAT      295    301       2-8.
FT   DOMAIN      306    347       P-TYPE 2.
FT   DOMAIN      353    394       P-TYPE 3.
FT   DOMAIN      402    522       12 X APPROXIMATE TANDEM REPEATS, THR-
FT                                RICH.
FT   REPEAT      402    411       3-1.
FT   REPEAT      412    419       3-2.
FT   REPEAT      420    431       3-3.
FT   REPEAT      432    443       3-4.
FT   REPEAT      444    453       3-5.
FT   REPEAT      454    460       3-6.
FT   REPEAT      461    472       3-7.
FT   REPEAT      473    479       3-8.
FT   REPEAT      480    491       3-9.
FT   REPEAT      492    498       3-10.
FT   REPEAT      499    515       3-11.
FT   REPEAT      516    522       3-12.
FT   DOMAIN      525    566       P-TYPE 4.
FT   DOMAIN      572    613       P-TYPE 5.
FT   DOMAIN      620    661       P-TYPE 6.
FT   DISULFID    162    188       BY SIMILARITY.
FT   DISULFID    172    187       BY SIMILARITY.
FT   DISULFID    182    199       BY SIMILARITY.
FT   DISULFID    307    333       BY SIMILARITY.
FT   DISULFID    317    332       BY SIMILARITY.
FT   DISULFID    327    344       BY SIMILARITY.
FT   DISULFID    354    380       BY SIMILARITY.
FT   DISULFID    364    379       BY SIMILARITY.
FT   DISULFID    374    391       BY SIMILARITY.
FT   DISULFID    526    552       BY SIMILARITY.
FT   DISULFID    536    551       BY SIMILARITY.
FT   DISULFID    546    563       BY SIMILARITY.
FT   DISULFID    573    599       BY SIMILARITY.
FT   DISULFID    583    598       BY SIMILARITY.
FT   DISULFID    593    610       BY SIMILARITY.
FT   DISULFID    621    647       BY SIMILARITY.
FT   DISULFID    631    646       BY SIMILARITY.
FT   DISULFID    641    658       BY SIMILARITY.
FT   VARSPLIC    240    259       Missing (in isoform 5 and isoform 6).
FT                                /FTId=VSP_004646.
FT   VARSPLIC    250    259       Missing (in isoform 4 and isoform 7).
FT                                /FTId=VSP_004647.
FT   VARSPLIC    276    294       Missing (in isoform 4 and isoform 6).
FT                                /FTId=VSP_004648.
FT   VARSPLIC    278    278       Missing (in isoform 5).
FT                                /FTId=VSP_004649.
FT   VARSPLIC    306    350       Missing (in isoform 2 and isoform 5).
FT                                /FTId=VSP_004650.
FT   VARSPLIC    420    498       Missing (in isoform 3).
FT                                /FTId=VSP_004651.
FT   VARIANT     276    276       K -> E.
FT   VARIANT     354    354       C -> R.
FT   VARIANT     415    415       T -> A.
SQ   SEQUENCE   662 AA;  67774 MW;  F085277F1ED2FD40 CRC64;

Alignment Scores:
Pred. No.:              0.0029          Length:         662
Score:                  192.00          Matches:        133
Percent Similarity:     32.76%          Conservative:   78
Best Local Similarity:  20.65%          Mismatches:     237
Query Match:            4.73%           Indels:         196
DB:                     1               Gaps:           30

US-09-914-958B-35 (1-2177) x MUC1_XENLA (1-662)

Qy          7 GCTCGAGGCCAGCCC------GGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              |||   |||   |||       ||||||              ||||||||||||
Db         95 AlaAlaGlyAlaProThrThrAlaThrGlyLysAlaProAlaThrAlaAlaAlaProVal 114

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGA 120
                |||||||||   :::   |||         ||||||         :::
Db        115 ProThrThrAlaAlaSer---LysAlaProThrThrAlaAlaAlaAlaThrHisSerThr 133

Qy        121 GGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGA----------------------- 156
                    |||   |||   |||
Db        134 AlaAlaAlaAlaAlaProThrThrAlaAlaSerAlaAlaLysSerLysGluArgSerThr 153

Qy        156 ------------------------------------------------------------ 156

Db        154 SerSerSerSerGluGluGluHisCysHisValLysProSerLysArgGluMetCysGly 173

Qy        157 ----------------------TGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCA 192
                                    |||   :::     ||||||     |||   |||
Db        174 SerLysGlyIleThrLysLysGlnCysLysLysLysAsnCysCys------PheAspPro 191
```

```
Qy      193 TCACTGAGGGGT-----------------------CCCAGCGGGCTGAAC--------- 219
Db      192 LysGlyHisGlyGlyIleHisCysPheHisArgLysProLysGlyHisSerHisGluGlu 211

Qy      220 ------CCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATC 273
Db      212 HisThrThrThrThrThrLysAlaProThrThrIleGlnIleAlaThrThrThrThrThr 231

Qy      274 CCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGA 333
Db      232 ProThrThrThrThrThrThrLysAlaThrProThrThrThrThrThrThrLysAla 251

Qy      334 TCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGC 393
Db      252 Thr----------------------------------------------------ProThrThr 255

Qy      394 GGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGC 453
Db      256 ThrThrThrThrLysAlaThrThrThrThrThrThrProThrThrThrThrThrThrThr 275

Qy      454 AGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCT 513
Db      276 LysAlaThrThrThrProThrThrThrThrThrThrProThrThrThrThrThrLys 295

Qy      514 ACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA--------- 564
Db      296 AlaThrThrThrThrThrThrSerGlyGluCysLysMetGluProSerLysArgGlu 315

Qy      565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
Db      316 AspCysGlyTyrSerGlyIleThrGluSerGlnCys---------------ArgThrLys 330

Qy      625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
Db      331 GlyCysCysPheAspSerSerIleProGlnThrLysTrpCysPheTyrThrLeu------ 348

Qy      685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
Db      349 ------SerGlnValAlaAspCysLysValGluProSer------------------Gln 360

Qy      745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
Db      361 ArgValAspCysGlyPheArgGlyIleThrAlaAspGlnCysArgGlnLysAsnCysCys 380

Qy      805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
Db      381 Phe---------------AspSerSerIleSerGlyThrLys---------TrpCysPhe 392

Qy      865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
Db      393 TyrSerThrSerGlnValAlaAlaThrLysThrThrThrThrProThrThrThrThrThr 412

Qy      925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
Db      413 ProThrThrThrThrThrThrLysAlaThrThrThrThrProThrThrThrThrThrThr 432

Qy      985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
Db      433 ProThrThrThr------------------ThrThrThrThrThrThrThrLysAlaThr 446

Qy     1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
Db      447 ThrThrThrProThrThrThrThrProThrThrThrThrThrLys--------------- 461

Qy     1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
Db      462 ------AlaThrThrThrThrPro-------------------------ThrThrThr 470

Qy     1165 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 1224
Db      471 ThrThrThrProThrThrThrThrLysAlaThrThrThrThrProThrThr---Thr 489

Qy     1225 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 1284
Db      490 ThrThrThrProThrThrThrThrThrLysAlaThrThrThrThrProThrThrThrThr 509

Qy     1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGG-------------------------- 1317
Db      510 ThrThrThrThrThrThrLysAlaThrThrThrThrThrSerGlyGluCysLysMetGlu 529

Qy     1318 ---AGCCTGAGGGCC------GGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACG 1368
Db      530 ProSerLysArgAlaAspCysGlyTyrProGlyIle---------------ThrGluSer 544

Qy     1369 GGATGCTGGACCTCATCTTGTCCCATGGAGAGT-------------------------- 1401
Db      545 GlnCysArgSerLysGlyCysCysPheAspSerSerIleProGlnThrLysTrpCysPhe 564

Qy     1402 ---------CCATGGCTCAGCCGCTGT---------CCGTCTTCC--------------- 1428
Db      565 TyrSerLeuProGlnValAlaAspCysLysValAlaProSerSerArgValAspCysGly 584

Qy     1429 ---GGGGCAATCAGGGCT--------------------------------TCAACAACA 1452
Db      585 PheGlyGlyIleThrAlaAspGlnCysArgGlnArgAsnCysCysPheAspSerSerIle 604

Qy     1453 ACTGGCTGCGAGTGGTGC--------------------CACGCACCCGGTTTGGGGCCT 1491
Db      605 SerGlyThrLysTrpCysPheTyrSerThrSerGlnGlyAsnAlaMetCysSerGlyPro 624

Qy     1492 TTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCA 1551
Db      625 ---ProThrLysArgArgAspCysGlyTyrPro---------------------GlyIle 636

Qy     1552 TCGACGGGGGCTCAGGCTACCTGTGTGAGA------------TGGAGC------------ 1587
Db      637 SerSer---------SerValCysIleAsnArgGlyCysCysTrpAspAsnSerValMet 653

Qy     1588 ------CCGTGG 1593
Db      654 AsnValProTrp 657
```

RESULT 10
PYGO_DROME
ID   PYGO_DROME   STANDARD;   PRT;   815 AA.
AC   Q9V9W8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pygopus protein (Gammy legs protein).
GN   PYGO OR GAM OR CG11518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21952490; PubMed=11955446;
RA   Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA   Murone M., Zuellig S., Basler K.;
RT   "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT   pygopus to the nuclear beta-catenin-TCF complex.";
RL   Cell 109:47-60(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Head;
RX   MEDLINE=22010046; PubMed=12015286;
RA   Parker D.S., Jemison J., Cadigan K.M.;
RT   "Pygopus, a nuclear PHD-finger protein required for wingless signaling
RT   in Drosophila.";
RL   Development 129:2565-2576(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;

```
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC   -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and
CC       thereby becomes part of the nuclear ARM/PAN complex.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Ubiquitous througout embryogenesis and larval
CC       development.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC       throughout development.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF457206; AAL91369.1; -.
DR   EMBL; AY075095; AAL79357.1; -.
DR   EMBL; AE003778; AAF57161.1; -.
DR   EMBL; AY058500; AAL13729.1; -.
DR   FlyBase; FBgn0043900; pygo.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0030528; F:transcription regulator activity; IPI.
DR   GO; GO:0030177; P:positive regulation of Wnt receptor signali. . .; IPI.
DR   GO; GO:0007367; P:segment polarity determination; IMP.
DR   GO; GO:0016055; P:Wnt receptor signaling pathway; IGI.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Nuclear protein; Trans-acting factor; Wnt signaling pathway; Zinc;
KW   Metal-binding; Zinc-finger; Segmentation polarity protein.
FT   DOMAIN       39     45       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     747    805       PHD-TYPE.
FT   DOMAIN       48     65       ALA-RICH.
FT   DOMAIN      123    749       ASN/GLY/HIS/MET/PRO-RICH.
FT   CONFLICT    393    393       S -> P (IN REF. 1).
SQ   SEQUENCE   815 AA;  80493 MW;  369FD5A5D34BC136 CRC64;

Alignment Scores:
Pred. No.:                0.00307        Length:          815
Score:                    192.00         Matches:         171
Percent Similarity:       30.83%         Conservative:    55
Best Local Similarity:    23.33%         Mismatches:      265
Query Match:              4.64%          Indels:          243
DB:                       1              Gaps:            38

US-09-914-958B-35 (1-2177) x PYGO_DROME (1-815)

Qy       1827 CGAATGGGAACTGGATGCATTCATTGGTGTCCATGCAATGGCCATTTTCCTGCTGGGAGA 1768
              :::||||||   |||                |||      |||||||||      |||
Db        121 GlnMetGlyGlyGly--------------ProGlyGlyGlyHisPheGlyGlyGlyGly 135

Qy       1767 ATCCTTGGCCAC-------------------------------------------ACTCCA 1750
                 ||||||                                              |||
Db        136 ProGlyGlyHisProHisTrpGluAspArgMetGlyMetGlyGlyGlyProProProPro 155

Qy       1749 GTGGGGCTGGGTCCTGAAGTGTGTCCT---CATCCCGGGGGT---AGAGGATCTCCAGCA 1696
                 :::   |||   :::   |||   ||||||||||||      |||   |||
Db        156 ProHisMetHisProHisMetHisProHisHisProGlyGlyProMetGlyHisPro--- 174

Qy       1695 CTGAGTTCATCTCCCCGCTGGCCACGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCA 1636
                        |||      |||      |||   |||                   |||
Db        175 ------HisGlyProHis---ProHisMetGlyGlyProProProMetArgGlyMetSer 191

Qy       1635 CCT------CCACACTGCTGGCTTCATCCTTCCCCAGGCCAAAGTGTGCCACGG------ 1588
                 |||      ||||||      :::   |||   ||||||      :::|||
Db        192 ProMetHisProHis---GlnMetGlyProGlyProGlyValGlyLeuProProHisMet 210

Qy       1588 ------------------------------------------------------------ 1588

Db        211 AsnHisGlyArgProGlyGlyProGlyGlyProGlyGlyProValProMetGlySerPro 230

Qy       1588 ------------------------------------------------------------ 1588

Db        231 MetGlyGlyIleAlaGlyMetGlyGlyMetSerProMetGlyGlyMetGlyGlyProSer 250

Qy       1587 ---GCTCCATCTCAC------ACAGGTAGCCTGAGCCCCCGTCGATGATCCTCAGGTGGG 1537
                  :::|||   |||          |||   |||||||||              ||||||
Db        251 IleSerProHisHisMetGlyMetGlyGlyLeuSerPro---------MetGlyGlyGly 267

Qy       1536 CCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTCCCCTGGCAAAGGCCCCAAACCGGG 1477
                 |||:::          ||||||          ||||||      ||||||
Db        268 ProAsnGlyProAsnProArgAlaMetGln-GlySerProMetGlyGlyProGlyGlnAs 287

Qy       1476 TGCGTGGCACCACTCGCAGCCAGTT-----------------GTTGTTGAAGCCCTGAT 1435
                  :::      :::|||   |||:::                      :::   :::   ::
Db        287 nSerProMetAsnSerLeuProMetGlySerProMetGlyAsnProIleGlySerProLe 307

Qy       1434 TGCCCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGATGAGGTCCA 1375
                 :   |||      |||          :::      :::   |||      :::   ||
Db        307 uGlyProProSerGlyProGlyProGlyAsnProGlyAsnThrGlyGlyProGlnGlnGl 327

Qy       1374 GCATCCCGTCTCCGTCGAAGTCGGTCACCACACCCCCTGTGCCCCGGCCCTCAGGCTCCA 1315
                 |          |||                |||      |||      :::|||
```

```
Db        327 nGln-GlnGlnPro------------------ProGlnProProMetAsnAsnGlyGlnM 341
Qy       1314 AGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGATGACGC 1255
Db        341 etGlyProProProLeuHisSerProLeuGlyAsnGlyProThrGly-HisGlySerHis 360
Qy       1254 GGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGAAGATCTCCAGCT 1195
Db        361 MetProGlyGlyProIleProGlyProGly-----------------------ProGly 372
Qy       1194 CCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTGCGGACAGGGGAGGGCATGGAGAACT 1135
Db        373 ProGlyGlyLeu---------------------ValGlyProGly--------------- 380
Qy       1134 TGGGTGAGGCGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGCAGATAGAGGC 1075
Db        381 ------------------GlyIleSerProAlaHisGlyAsnAsnProGlyGlySerGly 394
Qy       1074 ------------------------GGTGGGGGCCATTCCAGTTGCCATAGACGATGTCCA 1039
Db        395 AsnAsnMetLeuGlyGlyAsnProGlyGlyGlyAsnSerAsn---AsnAsnGlySerAsn 413
Qy       1038 CTTTGCCATCACGGTTGAAGTCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGG---- 983
Db        414 ThrSerAsnAlaSerAsnAsnAsnGlnAsnProHisLeuSerProAlaAlaGlyArgLeu 433
Qy        982 ---------------------TCGTCCACACCAGCACTGGCCGCA---GCGTCCACAAAG 947
Db        434 GlyValProThrSerMetGlnSerAsnGlyProSerValSerSerValAlaSerSerSer 453
Qy        946 GTGCCATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTGTCGCAGAAGATA 887
Db        454 ValProSerProAla------ThrProThrLeuThrProThrSerThrAlaThrSerMet 471
Qy        886 TCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCCCCTGTATATTTG 827
Db        472 SerThrSerVal--------------ProThrSerSerProAlaProProAla------ 484
Qy        826 CTGACCCCAGCCTCAGCAGCCACATCTCTG------AGCGCCAGAATGCCCCGGGAGAGG 773
Db        485 MetSerProHisHisSerLeuAsnSerAlaGlyProSer-ProGlyMetProAsnSerGl 504
Qy        772 TCACTGGCCT---CAGGGTCCATTTCAATGAGGGCATCAGGGCCCACATTACCGTAGGCG 716
Db        504 yProSerProLeuGlnSerProAla------GlyProAsnGlyProAsnAsnAsnAsnSe 522
Qy        715 TAATTGGCAATGTAGATAGAGTAGCGTCCAGAGCCCTTTC------------------- 676
Db        522 rAsn-----------------AsnAsnAsnGlyProMetMetGlyGlnMetIleProAs 536
Qy        675 ----TGTCCACACAGGCCACAGAGCGTCCGGCAAAGAGGCTGGCCACACCACGGGCCA-- 622
Db        536 nAlaValProMetGlnHisGlnGlnHisMetGlyGlyGlyProProGlyHisGlyProGl 556
Qy        622 ------------------------------------------------------------ 622
Db        556 yProMetProGlyMetGlyMetAsnGlnMetLeuProProGlnGlnProSerHisLeuGl 576
Qy        621 ---------------------CGTTGACCTCATCGCTCAGGATGTCTTCCCACCGGTTA 584
Db        576 yProProHisProAsnMetMetAsnHisProHisHisProHisHisHisProGlyGly-- 595
Qy        583 TTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGGCATTATTGGTG 524
Db        596 --------------------------------ProProProHisMetMetGlyGlyPr 604
Qy        523 TTGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGTCGATGTCGCAGGCTGTGACCCCAATG 464
Db        604 oGlyMetHisGlyGlyProAlaGlyMetProProHisMetGlyGly---GlyProAsnPr 623
Qy        463 GCGTTCCCCTGCCGGTCCCGCAGCGCGTAGTAGGGTGAGCTGCGCTCATCGACCGCGATG 404
Db        623 oHisMetMetGlyGlyProHisGlyAsnAlaGlyProHisMetGlyHisGlyHisMetGl 643

Qy        403 TTCACCAGCCGCTTCTGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCATTGTACCCC 344
Db        643 yGlyValProGlyProGlyProGly-----------ProGlyGlyMetAsn---GlyPr 658
Qy        343 GCCACGACGATCTCAAAGTCCCCATCATGGT---CCACATCAGTAACTGCCACACCATAG 287
Db        658 oProHisProHisMetSerProHisHisGlyHisProHisHis---------HisHisAs 675
Qy        286 TTGAGCTGGGTGGGATTACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTG 227
Db        675 n-ProMetGlyGly---------------ProGlyProAsnMetPheGlyGlyGlyGlyG 690
Qy        226 AACATGGGTTCAGCCCGCTGGGACCCCTCAGTGATGGGCAGAAACCAGAGCAGCAGCAGG 167
Db        690 lyGlyProMetGlyProGlyGlyProMetGly---------------------------- 700
Qy        166 AACGGTAACATCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCGGCC 107
Db        701 --------AsnMetGlyProMetGlyGlyGlyProMetGlyGlyProMetGlyValGly- 717
Qy        106 CCGCCGCCTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCGCCT------------------ 65
Db        718 ProLysProMetThrMetGlyGlyGlyLysMetTyrProProGlyGlnProMetValPhe 737
Qy         64 ---------CTGCCGCCGGCGCCGCCG 47
Db        738 AsnProGlnAsnProAsnAlaProPro 746
```

RESULT 11
P121_RAT
ID   P121_RAT   STANDARD;   PRT; 1199 AA.
AC   P52591;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE   of 121 kDa) (P145).
GN   POM121 OR NUP121.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=93328754; PubMed=8335683;
RA   Hallberg E., Wozniak R.W., Blobel G.;
RT   "An integral membrane protein of the pore membrane domain of the
RT   nuclear envelope contains a nucleoporin-like region.";
RL   J. Cell Biol. 122:513-521(1993).
CC   -!- FUNCTION: Essential component of the nuclear pore complex. The
CC   repeat-containing domain may be involved in anchoring components
CC   of the pore complex to the pore membrane.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC   MEMBRANE.
CC   -!- DOMAIN: Contains F-X-F-G repeats.
CC   -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC   SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC   NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC   ---

CC   ---
DR   EMBL; Z21513; CAA79725.1; -.
DR   EMBL; Z21514; CAA79726.1; -.
DR   PIR; A40670; A40670.
DR   InterPro; IPR004325; Nucleoporin_FG.

```
DR   Pfam; PF03093; Nucleoporin_FG; 8.
KW   Nuclear protein; Transport; Transmembrane; Repeat.
FT   DOMAIN        1     56       CISTERNAL SIDE (POTENTIAL).
FT   TRANSMEM     57     77       POTENTIAL.
FT   DOMAIN       76   1199       PORE SIDE (POTENTIAL).
FT   DOMAIN        4      8       POLY-ALA.
FT   DOMAIN       53     58       POLY-ALA.
FT   DOMAIN      438    441       POLY-SER.
FT   DOMAIN      497    500       POLY-PRO.
FT   DOMAIN      684    687       POLY-SER.
FT   DOMAIN     1023   1026       POLY-ALA.
FT   DOMAIN     1033   1038       POLY-SER.
FT   DOMAIN     1058   1063       POLY-GLY.
SQ   SEQUENCE   1199 AA;  120784 MW;  6DC4451B91D5B907 CRC64;

Alignment Scores:
Pred. No.:                0.00364        Length:          1199
Score:                    191.50         Matches:         190
Percent Similarity:       32.86%         Conservative:    63
Best Local Similarity:    24.68%         Mismatches:      307
Query Match:              4.72%          Indels:          210
DB:                       1              Gaps:            35

US-09-914-958B-35 (1-2177) x P121_RAT (1-1199)

Qy         20 CCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGG------------CAGAGG 67
Db        471 ProThrProGlySerSerGlyGlnArgLysArgLysIleGlnLeuLeuProSerArgArg 490

Qy         68 CGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGATGGC 127
Db        491 -GlyAspGlnLeuThrLeuProProPro------------ProGluLeuGlyTyrSerIl 506

Qy        128 TCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTT-- 185
Db        506 eThrAlaGluAspLeuAspMetGluArgArgAlaSer---------LeuGlnTrpPheAs 523

Qy        186 ---------------------------------CTGCCCATCACTGAGGGGTCCCAGCG 211
Db        523 nLysValLeuGluAspLysThrAspAspAlaSerThrProAlaThrAspThrSerProAl 543

Qy        212 GGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCT----------------- 254
Db        543 aThrSerProProPheThrLeuThrLeuProThrValGlyProAlaAlaSerProAlaSe 563

Qy        255 ----CCTGACTATGACAGTAATCCC---------------ACCCAGCTCAACTATGGTGT 295
Db        563 rLeuProAlaProSerSerAsnProLeuLeuGluSerLeuLysLysMetGlnGluSerPr 583

Qy        296 GGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGG 355
Db        583 oAlaProSer------SerSerGluProProGluAlaAlaThrValAlaAlaProSerPr 601

Qy        356 ACCCAAC---------------------CTGGTTCTGAAGTATGACCGGGCCCAGAAGCG 394
Db        601 oProLysThrProSerLeuLeuAlaProLeuValSer-ProLeuThrGlyPro------- 618

Qy        395 GCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGC------------T 442
Db        619 --------LeuAlaSerThrSerSerAspSerLysProThrThrThrPheLeuGlyLeuA 636

Qy        443 GCGGGACCGGCAGGGGAACGCCATTGGGG---------------------TCACAGCCTG 481
Db        636 laSerAlaSerSerAlaThrProLeuThrAspThrLysAlaProGlyValSerGlnAlaG 656

Qy        482 CGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTC 541
Db        656 lnLeuCysValSerThrProAlaAlaThrAlaProSerProThrProAlaSerThrLeuP 676

Qy        542 GGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACAT 601
Db        676 heGlyMetLeuSerProProAlaSerSerSerSerLeuAlaThrProGlyProAlaCysA 696

Qy        602 CCTGAGCGATG---AGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
Db        696 laSerProMetPheLysProIlePheProAlaThr---ProLysSerGluSerAspAsnP 715

Qy        659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACA---TTGCCAATTA 715
Db        715 roLeuProThrSerSerSerAlaAlaThrThrThrProAlaSerThrAlaLeuProThrT 735

Qy        716 CGCCTACGGTAATG----------------------------------------TGGGCCC 736
Db        735 hrAlaThrAlaThrAlaAlaHisThrPheLysProIlePheGluSerValGluProPheAlaA 755

Qy        737 TGATGCCC--------------TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCG 781
Db        755 laMetProLeuSerProProPheSerLeuLysGlnThrThrAlaProAlaThrThrAlaA 775

Qy        782 GGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAAT---------A 832
Db        775 laThr---------SerAlaProLeuLeuThrGlyLeuGlyThrAlaThrSerThrValA 792

Qy        833 TACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA---TCCTCAGCAGCAGTGCCTCGGATAT 889
Db        792 laThrGlyThrThrAlaSerAlaSerLysProValPheGlyPheGlyValThrThrAlaA 812

Qy        890 CTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTT 949
Db        812 laSerThrAlaSerThrIleAlaSerThrSerGlnSerIleLeuPheGlyGlyAlaPro- 831

Qy        950 TGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGC 1009
Db        832 ----------------------------------ProValThrAlaSerSerSerAlaP 840

Qy       1010 CCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCC 1069
Db        840 roAlaLeuAlaSerIlePhe-----------------GlnPheGlyLysProLeuAlaP 854

Qy       1070 CCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTC 1129
Db        854 roAlaAlaSerVal----------------------------AlaGlyThrSerPheS 864

Qy       1130 ACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCG------------- 1176
Db        864 erGlnSerLeuAlaSerSerAlaGlnThrAlaAlaSerAsnSerSerGlyGlyPheSerG 884

Qy       1177 -----------ACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTA 1225
Db        884 lyPheGlyGlyThrLeuThr--------------------ThrSerThrSerAlaProA 897

Qy       1226 CCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACG---GAGACCC 1282
Db        897 laThrThrSerGlnProThrLeuThrPheSerAsnThrValThrProThrPheAsnIleP 917

Qy       1283 CCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGG 1342
Db        917 roPheSerAlaSerAlaLysProAlaLeuProThrTyrProGlyAlaAsnSerGlnProT 937

Qy       1343 TGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTC 1402
Db        937 hrPheGlyAlaThrAspGlyAlaThr-----------------------------LysP 947

Qy       1403 CATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCG 1462
Db        947 roAlaLeuAla------ProSerPheGly---------SerSerPheThrPheGlyAsnS 962

Qy       1463 AGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCA------------------------- 1497
Db        962 erValAlaSerAlaProSerAlaAlaProAlaProAlaAlaPheGlyGlyAlaAlaGlnP 982

Qy       1498 --------GGGGAGCTAAGGTCGTGCTCT-----------ACACCAAGAAGAGTGGGGC 1537
Db        982 roAlaPheGlyGlyLeuLysAlaSerAlaSerThrPheGlyThrProAlaSerThrGlnP 1002

Qy       1538 CCACCTGAGGATCATCGACG---------------------------------------- 1557
```

```
              ||      ||||||:::|||
Db       1002 roAlaPheGlySerThrThrSerValPheSerPheGlySerAlaThrThrSerGlyPheG 1022

Qy       1558 -----GGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAA 1612
                    |||      |||          |||          :::      :::|||
Db       1022 lyAlaAlaAlaAlaThrThrGlnThrThrHisSerGlySerSerSerSerLeuPheGlyS 1042

Qy       1613 GGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGT 1672
                    |||:::   :::          |||      |||
Db       1042 erSerThrProSerProPheThrPheGlyGlySerAlaAla------------------- 1055

Qy       1673 GGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACT 1732
                |||||||||                  |||:::|||||||||    |||   |||
Db       1056 --ProAlaGlyGly---GlyGlyPheGlyLeuSerAlaThrProGlyThr-GlySerThr 1073

Qy       1733 TCAGGACCCAGCCCCACTGGAGTGTGGCCAA---------------------------- 1763
                ||||||          |||    ||||||
Db       1074 SerGlyThrPheSerPheGly-SerGlyGlnSerGlyThrThrGlyThrThrThrSerPh 1093

Qy       1764 ---------------------------GGATTCTCCCAGCAGGAAAATGGCCATTGCAT 1795
                                        ||| :::||||||:::          |||:
Db       1093 eGlyGlySerLeuSerGlnAsnThrLeuGly-AlaProSerGlnSerSerProPheAlaP 1113

Qy       1796 GGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGC---CCGTATG 1852
                :::::      :::          |||                  ||||||   |||
Db       1113 heSerValGlySerThrProGluSerLysProValPheGlyGlyThrSerThrProThrP 1133

Qy       1853 TGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGA 1912
                          |||   |||  |||   |||   |||::::::      :
Db       1133 heGlyGlnSerAlaProAlaProGlyValGlyThrThrGlySerSerLeuSerPheGlyA 1153

Qy       1913 GCCCAACGAGGATGGCACAG 1932
                ::|||:::      ||||||
Db       1153 laProSerThrProAlaGln 1159

RESULT 12
MUC1_MESAU
ID   MUC1_MESAU     STANDARD;      PRT;   676 AA.
AC   Q60528;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 1 precursor.
GN   MUC1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tracheal epithelium;
RX   MEDLINE=96326118; PubMed=8703480;
RA   Park H., Hyun S.W., Kim K.C.;
RT   "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
RT   cells in primary culture.";
RL   Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC   -!- SIMILARITY: Contains 1 SEA domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U36918; AAB53965.1; -.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat.
FT   SIGNAL        1    25       POTENTIAL.
FT   CHAIN        26   676       MUCIN 1.
FT   DOMAIN       26   582       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    583   603       POTENTIAL.
FT   DOMAIN      604   676       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      458   573       SEA.
FT   CARBOHYD    291   291       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    323   323       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    350   350       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    380   380       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400   400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413   413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435   435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479   479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    496   496       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536   536       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  67616 MW;  95F479B6EC5C3884 CRC64;

Alignment Scores:
Pred. No.:                0.00331       Length:          676
Score:                    191.00        Matches:         146
Percent Similarity:       37.16%        Conservative:    71
Best Local Similarity:    25.00%        Mismatches:      221
Query Match:              4.71%         Indels:          146
DB:                       1             Gaps:            30

US-09-914-958B-35 (1-2177) x MUC1_MESAU (1-676)

Qy        229 CTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACT 288
              :::::::::|||           ||||||      |||      |||:::|||
Db          5 IleArgAlaProPheLeuLeuThrLeuLeuLeuAlaLeuValThrAspProAsnSerVal 24

Qy        289 ATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGT 348
              :::   |||                        |||      ||||||
Db         25 AlaLeuSerGln--------------------AspThrSerSerSerSerThrLeuAsn 37

Qy        349 ACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCG 408
              |||                       |||          ||||||
Db         38 Thr-------------------------ThrProValHisSerGlySerSerAlaPro 48

Qy        409 CGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGG---ACCGGCAGGGGAACGCCA 465
              :::   ||||||           |||      |||   :::|||          |||
Db         49 AlaThrSerSerAlaValAspSerAlaThrThrProGlyHisSerGlySerSerAlaPro 68

Qy        466 TTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACA 525
                 |||           |||      |||   |||         :::|||:::
Db         69 ProThrSerSerAlaValAsnSerAlaThrThrProGlyHisSerGlySerSerAlaPro 88

Qy        526 CCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATA 585
              |||      :::          ||||||          |||||||||
Db         89 ProThrSerSerAlaValAsnSerAlaThrThrProValHisSerGlySerSerAlaPro 108

Qy        586 ACCGGTGGGAAGACATCCTGAGCGATGAGGTCA---ACGTGGCCCGTGGTGTGGCCAGCC 642
                          ||||||   |||:::   |||   |||   ||||||      :::
Db        109 Val---------ThrSerSerAlaValAsnSerAlaThrThrProValHisSerGlySer 125

Qy        643 TCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTA--- 699
              |||   |||                    |||      |||:::::::::
Db        126 SerAlaProPro-----------------ThrSerSerAlaValAsnSerAlaThrThr 139

Qy        700 --------------TCTACATTGCCAATTACGCCTACG---GTAATGTGGGCCCTGATG 741
                          |||:::   |||:::|||   :::   |||      |||
Db        140 ProValHisSerGlySerSerAlaProValThrSerSerAlaValAsnSerAlaThrThr 159

Qy        742 CCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAG 801
              |||           :::      ||||||||||
Db        160 ProValHisSerGlySerSerAlaProValThrSerSerAla------------------ 173
```

```
Qy       802 ATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCC 861
Db       174 --------------ValAspSerAlaThrThrProValHisSerGlySerSerAlaPro 188

Qy       862 CCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCC 921
Db       189 ProThrSerSerAlaVal---------AsnSerAlaThrThrProVal------------ 201

Qy       922 TTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACG 981
Db       202 ------HisSerGlySerSerAlaProVal---ThrSerSerAlaValAsnSerAlaThr 218

Qy       982 ACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGG 1041
Db       219 ThrProValHisSerGlySerSerAlaPro---ValThrSerSerAlaValAsnSerAla 237

Qy      1042 ACATCGTCTATGGCAACTGGAATGGCC------CCCACCGCCTCTATCTGCAAATGAGCA 1095
Db       238 ThrThrProValHisSerGlySerSerAlaProProThrSerSerValValAsn------ 255

Qy      1096 CCCATGGGAAGGTCCGCTTCCGGGACATCGCCT---CACCCAAGTTCTCCATGCCCTCCC 1152
Db       256 ------------------SerAlaThrThrProValHisSerGlySerSerAlaProPro 269

Qy      1153 CTGTCCGCACGG---------TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGA 1203
Db       270 ThrSerSerAlaValAsnLeuAlaThrThrProValHisSerGlySerSerThrProAla 289

Qy      1204 TCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCC 1263
Db       290 ThrAsnSerThrThrAspSerAlaThrThrProValPro---------ProGlySerSer 306

Qy      1264 GTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTG 1323
Db       307 MetGlnThrThrGluAlaIleSerGlySerAlaAsnThrProIleHisAsnGlySerLeu 326

Qy      1324 AGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACG------GAGACGGGATGCTGG 1377
Db       327 ValProThrThrSerSerAlaLeuValProThrThrSerAlaAlaHisSerGlyAlaSer 346

Qy      1378 ACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATC 1437
Db       347 AlaMetThrAsnSerSerGluSerAsp---LeuAlaThrThrProIleAspSerGlyThr 365

Qy      1438 AGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCA 1497
Db       366 SerIleSerThrThr------------------LysAlaPro------------------ 373

Qy      1498 GGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGG---------CCCACCTGAGGA 1548
Db       374 ---------------AlaThrThrProValHisAsnGlySerLeuValProThrThrSer 388

Qy      1549 TCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGG 1608
Db       389 SerValLeuGlySerAlaThrThrLeuIleHis---AsnAspThrSerThrMet------ 405

Qy      1609 GGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGA 1668
Db       406 ------------------------------------------------------AlaThr 407

Qy      1669 ACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACA 1728
Db       408 ThrThrProValGlyAsnGlyThrGln---------SerSerValProSer--------- 421

Qy      1729 CACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCC 1788
Db       422 ----------------------------------------------------ArgHisPro 424

Qy      1789 ATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCG 1848
Db       425 ValThrProThrProProAlaValSerSerAsnSerThrIleAlaLeuSerThr------ 442

Qy      1849 TATGTGTCAACA---------CCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCA 1899

              ||| |||||| ||| :::::: ||| ||| ::: :::
Db       443 TyrTyrSerThrAlaLeuSerProAlaPheSerSerHisAlaAlaPro---GlnValSer 461

Qy      1900 GTCGGGGCTACG 1911
             |||||| :::
Db       462 ValGlyValSer 465
```

RESULT 13
DRPL_HUMAN
ID   DRPL_HUMAN   STANDARD;   PRT; 1185 AA.
AC   P54259; Q99495; Q99621; Q9UEK7;
DT   01-OCT-1996 (Rel. 34, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN   DRPLA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Cerebellum;
RX   MEDLINE=95144175; PubMed=7842016;
RA   Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA   Inoue T., Yamada M.;
RT   "Structure and expression of the gene responsible for the triplet
RT   repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";
RL   Nat. Genet. 8:177-182(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96026098; PubMed=7485154;
RA   Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Tsuji S.;
RT   "Molecular cloning of a full-length cDNA for dentatorubral-
RT   pallidoluysian atrophy and regional expressions of the expand alleles
RT   in the CNS.";
RL   Am. J. Hum. Genet. 57:1050-1060(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96262314; PubMed=8965642;
RA   Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA   Kidwai A.S., Ashworth R.G., Ross C.A.;
RT   "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT   brain.";
RL   Brain Res. Mol. Brain Res. 36:219-226(1996).
RN   [4]
RP   SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.
RC   TISSUE=Brain;
RX   MEDLINE=97228904; PubMed=9074930;
RA   Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;
RT   "Large-scale sequencing in human chromosome 12p13: experimental and
RT   computational gene structure determination.";
RL   Genome Res. 7:268-280(1997).
RN   [5]
RP   SEQUENCE OF 1-76 FROM N.A.
RX   MEDLINE=97005364; PubMed=8852663;
RA   Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y.,
RA   Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K.,
RA   Soerensen S.A., Potter N.T., Young S.R., Nakamura K., Nukina N.,
RA   Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T.,
RA   Kanazawa I., Yamada M.;
RT   "A unique origin and multistep process for the generation of expanded
RT   DRPLA triplet repeats.";
RL   Hum. Mol. Genet. 5:373-379(1996).
RN   [6]
RP   SEQUENCE OF 470-725 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=93315145; PubMed=8325628;
RA   Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT   "Novel triplet repeat containing genes in human brain: cloning,
RT   expression, and length polymorphisms.";
RL   Genomics 16:572-579(1993).

```
RN   [7]
RP   INTERACTION WITH WWP1 AND WWP2.
RX   MEDLINE=98313405; PubMed=9647693;
RA   Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan K., Kushi J.,
RA   Kaminsky Z., Kleiderlein J.J. Jr., Sharp A.H., Ross C.A.;
RT   "Atrophin-1, the DRPLA gene product, interacts with two families of
RT   WW domain-containing proteins.";
RL   Mol. Cell. Neurosci. 11:149-160(1998).
CC   -!- SUBUNIT: Interacts with WWP1 and WWP2.
CC   -!- TISSUE SPECIFICITY: Relatively high levels in the brain, ovary,
CC       testis and prostate. Lower levels in the liver, thymus and
CC       leukocytes.
CC   -!- POLYMORPHISM: The poly-Gln region of DRPLA is highly polymorphic
CC       (7 to 23 repeats) in the normal population and is expanded to
CC       about 49-75 repeats in DRPLA patients. Longer expansions result in
CC       earlier onset and more severe clinical manifestations of the
CC       disease.
CC   -!- DISEASE: Defects in DRPLA are the cause of dentatorubral-
CC       pallidoluysian atrophy (DRPLA) [MIM:125370], an autosomal dominant
CC       neurodegenerative disorder characterized by a loss of neurons in
CC       the dentate nucleus, rubrum, glogus pallidus and Luys'body.
CC       Clinical features are myoclonus epilepsy, dementia, and cerebellar
CC       ataxia. Onset of the disease occurs usually in the second decade
CC       of life and death in the fourth.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to several
CC       frameshifts.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D31840; BAA06626.1; -.
DR   EMBL; D38529; BAA07534.1; ALT_FRAME.
DR   EMBL; U23851; AAB50276.1; -.
DR   EMBL; U47924; AAB51321.1; -.
DR   EMBL; D63808; BAA23631.1; -.
DR   EMBL; L10377; -; NOT_ANNOTATED_CDS.
DR   PIR; G01763; G01763.
DR   Genew; HGNC:3033; DRPLA.
DR   MIM; 607462; -.
DR   MIM; 125370; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0007417; P:central nervous system development; TAS.
DR   InterPro; IPR002951; Atrophin.
DR   Pfam; PF03154; Atrophin-1; 2.
DR   PRINTS; PR01222; ATROPHIN.
KW   Triplet repeat expansion; Polymorphism; Epilepsy.
FT   DOMAIN       73     82       SER/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      302    305       POLY-PRO.
FT   DOMAIN      376    382       POLY-SER.
FT   DOMAIN      386    397       POLY-SER.
FT   DOMAIN      442    447       POLY-PRO.
FT   DOMAIN      479    483       POLY-HIS.
FT   DOMAIN      484    497       POLY-GLN.
FT   DOMAIN      504    507       POLY-PRO.
FT   DOMAIN      564    574       POLY-SER.
FT   DOMAIN      704    707       POLY-PRO.
FT   DOMAIN      802    815       ARG/ALA-RICH (MIXED CHARGE).
FT   DOMAIN      816    827       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      925    934       ARG/GLU-RICH (MIXED CHARGE).
FT   CONFLICT     94     94       MISSING (IN REF. 3).
FT   CONFLICT    333    333       H -> Y (IN REF. 1).
FT   CONFLICT    339    339       M -> I (IN REF. 3).
FT   CONFLICT    541    541       P -> T (IN REF. 6).
FT   CONFLICT   1028   1028       A -> G (IN REF. 1).
SQ   SEQUENCE   1185 AA;  124773 MW;  5135744CEE491C18 CRC64;
```

```
Alignment Scores:
Pred. No.:              0.00411     Length:          1185
Score:                  190.50      Matches:         165
Percent Similarity:     32.63%      Conservative:    52
Best Local Similarity:  24.81%      Mismatches:      231
Query Match:            4.61%       Indels:          218
DB:                     1           Gaps:            37

US-09-914-958B-35 (1-2177) x DRPL_HUMAN (1-1185)

Qy      1779 CCTGCTGGGAGAATCCTTGGCCACACTCCAGTGGGGCTGGGTCCTGAAGTGTGTCCTCAT 1720
             |||      |||:::           |||  |||      |||        ||||||
Db       201 ProThrSerArgMetPhe---GlnAlaProProGlyAlaProPro---------ProHis 216

Qy      1719 -----------CCCGGGGGTAGAGGATCTCCAGCACTGAGTTCATCTCCCCGCTGGCCA 1672
                       |||||||||  |||      ||||||      |||      |||
Db       217 ProGlnLeuTyrProGlyGlyThrGlyGly---ValLeuSerGlyProProMetGlyPro 235

Qy      1671 CGTTCCGGCTCACCATCTTGCCATCTGGCCACGTCACCTCCACACTGCTGGCTTCATCCT 1612
             :::  |||      :::      :::      |||                  ||||||
Db       236 LysGlyGlyGlyAlaAlaSerSerValGly---GlyProAsnGlyGlyLysGlnHisPro 254

Qy      1611 TCCCCAGGCCAAAGTGTGCCACGGGCTCCATCTCACACAGGTAGCCTGAGCCCCCGTCGA 1552
               |||      :::  |||  :::  |||  :::|||
Db       255 ProPro------ThrThrProIleSerValSerSerSerGly------------------ 266

Qy      1551 TGATCCTCAGGTGGGCCCCACTCTTCTTGGTGTAGAGCACGACCTTAGCTCCCCTGGCAA 1492
                :::||||||  |||                                  |||  :::
Db       267 ---AlaSerGlyAlaPro---------------------------------ProThrLys 274

Qy      1491 AGGCCCCAAACCGGGTGCGTGGCACCACTCGCAGCCAGTTGTTGTTGAAGCCCTGATTGC 1432
               |||  |||                                          |||
Db       275 ProProThrThr---------------------------------------ProValGly 281

Qy      1431 CCCGGAAGACGGACAGCGGCTGAGCCATGGACTCTCCATGGGACAAGATGAGGTCCAGCA 1372
               |||        :::|||  |||                             ||||||
Db       282 GlyGlyAsnLeuProSerAlaProPro-------------------------ProAla 292

Qy      1371 TCCCGTCTCCGTCGAAGTCGGTCACCACAC------CCCCTGTGCCCCGGCCCTCAGGCT 1318
               :::                |||||      |||      |||  |||  |||
Db       293 Asn-----------------PheProHisValThrProAsnLeuProProProProAla 306

Qy      1317 CCAAGGCGTCGCCGGGATTGAGCTCCTCGATGAGGGGGTCTCCGTGCTCTCTACGGATGA 1258
               |||        :::  :::          ||||||  |||
Db       307 LeuArgProLeuAsnAsnAlaSerAlaSerProProGlyLeuGlyAla------------ 322

Qy      1257 CGCGGAAGAGGCGGTTGGCTGAGGAGCTGCGGTAGGCAATGTTGTTGAAGAAGATCTCCA 1198
                                                                    |||
Db       323 ------------------------------------------------------GlnPro 324

Qy      1197 GCTCCTGGTCATTGTCAAAGTCGGCGGTGATGACCGTG-CGGACAGGGGAGGGCATGGAG 1139
               |||||||||     |||          |||      |||:::||||||
Db       325 LeuProGlyHisLeuProSer------------ProHisAlaMetGlyGlnGlyMetGly 340

Qy      1138 AACTTGGGTGAGGCGATGTCCCGGAAGCGGACCTTCCCATGGGTGCTCATTTGCAGATAG 1079
               |||          :::
Db       341 GlyLeuProProGlyProGluLys---------------------------------- 348

Qy      1078 AGGCGGTGGGGGCCATTCCAGTTGCCATAGACGATGTCCACTTTGCCATCACGGTTGAAG 1019
                  ||||||          |||  :::      :::||||||
Db       349 ---------GlyProThrLeuAlaPro---SerProHisSerLeuProProAlaSerSer 364

Qy      1018 TCAGCCAGGGCGACACCTCGCCCATGCTGGTGGGGGTCGTCCACACCAGCACTGGCCGCA 959
               ||||||  |||  |||          :::  ||||||:::  :::  ||||||
Db       365 SerAlaProAlaProProMetArgPheProTyrSerSerSerSerSerSerAlaAla 384

Qy       958 GCGTCCACAAAGGTGCCATCGCCCCGGTTGTGGAAAAGGAAGTTAGGCCCATTCTCATTG 899
               ||||||:::        |||                        ||||||
Db       385 AlaSerSerSerSerSerSerSer---------SerSerSerAlaSerProPheProAla 401

Qy       898 TCGCAGAAGATATCCGAGGCACTGCTGCTGAGGATGGGGCCCACGCTGACGCCTCGGCCC 839
               ||||||  :::                              |||      |||  |||
Db       402 SerGlnAlaLeuProSerTyr-----------------ProHisSerPhePro---Pro 414
```

```
Qy        838 CCTGTATATTTGCTGACCCCAGCCTCAGCAGCCACATCTCTGAGCGCCAGAATGCCCCGG 779
Db        415 ProThrSerLeuSerVal-SerAsnGlnProProLysTyrThrGlnProSerLeuProSe 434

Qy        778 GAGAGGTCACTGGCCTCAGGGTCCA--------------------------------TT 752
Db        434 rGlnAlaValTrpSerGlnGlyProProProProProProTyrGlyArgLeuLeuAlaAs 454

Qy        751 TCAATGAGGGCATCAGGGCCCACATTACCGTAGGCGTAATTGGCAATGTAGATAGAGTAG 692
Db        454 nSerAsnAlaHisProGlyProPheProProSerThrGlyAlaGln---------SerTh 471

Qy        691 CGTCCAGAGCCCTTTCTGTCCACAC--------------------AGGCCACAGAGCGT 653
Db        471 rAlaHisProProValSerThrHisHisHisHisGlnGlnGlnGlnGlnGlnGlnGl 491

Qy        652 CCGGCAAAGAGGCTGGCCACACCACGGGCCACGTTGACCTCATCGCTCAGGATGTCTTCC 593
Db        491 nGlnGlnGlnGlnGlnGlnHisHisGly--------------AsnSerGlyProProPr 506

Qy        592 CACCGGTTATTGCGGAACTTGAACAACTTGTCGGTGTACGTGGCCACCCCCGAGAAGGCA 533
Db        506 oProGly-----------------------------AlaPheProHisProLeuGlu-- 515

Qy        532 TTATTGGTGTTGAGGAAGTAGATCTCCTCCCGGCCGTCCCCGTCGATGTCGCAGGCTGTG 473
Db        516 ------------GlyGlySerSerHisHisAlaHisPro---------TyrAlaMetSe 528

Qy        472 ACCCCAATGGCGTTCC------------CCTGCCGGTCCCGCA------------GCGCG 437
Db        528 rProSerLeuGlySerLeuArgProTyrProProGlyProAlaHisLeuProProProHi 548

Qy        436 TAGTAGGGTGAGCTGCGCT------------CATCGACCGCGATGTTCACCAGCCGCTTC 389
Db        548 sSerGlnValSerTyrSerGlnAlaGlyProAsnGlyProProValSerSerSerSerAs 568

Qy        388 TGGGCCCGGTCATACTTCAGAACCAGGTTGGGTCCATTGTACCCCGCCACGACGATCTCA 329
Db        568 nSerSerSerSerThrSerGln---GlySerTyrProCysSerHisProSerProSerGl 587

Qy        328 AAGTCCCCATCATGGTCCACAT------CAGTAACTGCCACACCATAGTTGAGCTGGGTG 275
Db        587 nGlyProGlnGlyAlaProTyrProPheProProValPro-ThrValThrThrSerSerA 607

Qy        274 GGATT------ACTGTCATAGTCAGGAGGCAGAACTGAGTTGGTGACTGCAGTGAACATG 221
Db        607 laThrLeuSerThrValIleAlaThrValAlaSer--------------SerProAlaG 622

Qy        220 GGTTC---------AGCCCGCTGGGACCCCTCAGTGATGGGCAGAAA------------ 183
Db        622 lyTyrLysThrAlaSerProProGlyProProProTyrGlyLysArgAlaProSerProG 642

Qy        182 ---------------------------------------------CCAGAGCAGCAGCAGG 167
Db        642 lyAlaTyrLysThrAlaThrProProGlyTyrLysProGlySerProProSerPheArgT 662

Qy        166 AACGGTAACATCCTGGACATGCCGGGGTCAGCGCTCGGAGCCATCCTCCCGCTCTCGGCC 107
Db        662 hrGlyThrProProGlyTyrArgGlyThrSerProProAlaGlyProGlyThrPheLysP 682

Qy        106 CCGCCGCCTAGGGGCGTGGGAAGCGGGCGCTCGCTGCCGCCTCTGCCGCCGGCG------ 53
Db        682 roGlySer-ProThrValGlyProGly------------ProLeuProProAlaGlyPro 697

Qy         52 --------------------CCGCCGCCTGCTTGCTCCCAGCCCCGGTCCCGGGCTGGC 14
Db        698 SerGlyLeuProSerLeuProProProProAla-AlaProAlaSerGlyProProLeuSe 717

Qy         13 CTCGAGC 7
Db        717 rAlaThr 719
```

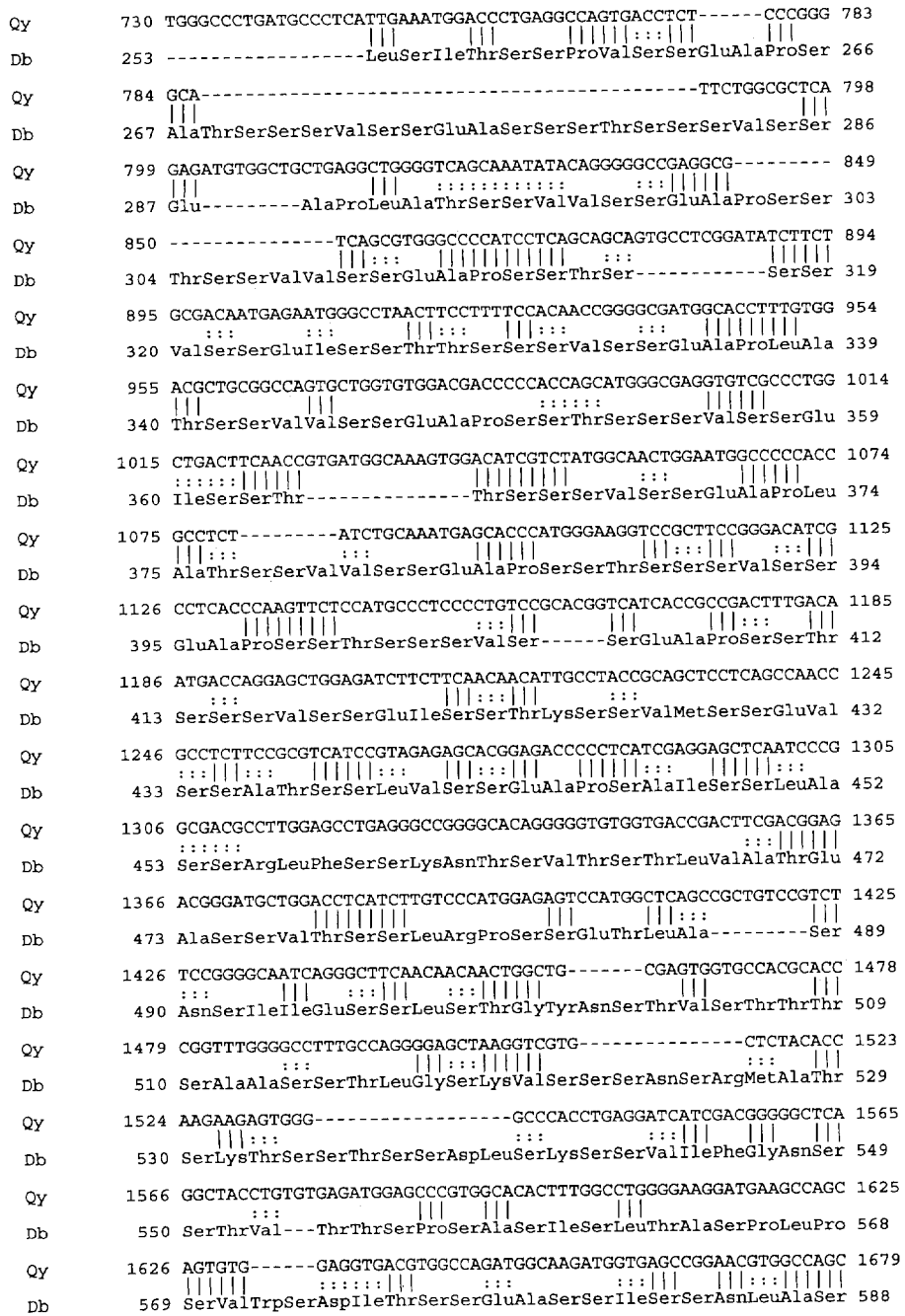
RESULT 14
YM96_YEAST
ID   YM96_YEAST   STANDARD;   PRT; 1140 AA.
AC   Q04893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN   YMR317W OR YM9924.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- DOMAIN: Contains many Ser/Thr-rich domain and repeats.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z54141; CAA90835.1; -.
DR   GermOnline; 142998; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Alignment Scores:

Pred. No.:	0.00672	Length:	1140
Score:	186.50	Matches:	144
Percent Similarity:	39.38%	Conservative:	86
Best Local Similarity:	24.66%	Mismatches:	242
Query Match:	4.60%	Indels:	113
DB:	1	Gaps:	25

US-09-914-958B-35 (1-2177) x YM96_YEAST (1-1140)

```
Qy        403 ACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACG 462
Db        140 SerSerSerAlaIleGlnThrSerSerProGlnThrSerSerSerAsnGlyGlyGlySer 159

Qy        463 ---------CCATTGGGGTCACAG--------------CCTGCGACATCGACGGGGACG 498
Db        160 SerSerGluProLeuGlyLysSerSerValLeuGluThrThrAlaSerSerSerAspThr 179

Qy        499 GCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACA 558
Db        180 ThrAlaValThrSerSerThrPheThrThrLeuThrAspValSerSerSerProLys--- 198

Qy        559 CCGACAAGTTGTTCAAGTTCCGCAATAACCGGT---GGGAAGACATCCTGAGCGATGAGG 615
Db        199 ---IleSerSerSerGlySerAlaValThrSerValGlyThrThrSerAspAlaSerLys 217

Qy        616 TCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC------GCTCTGTGGCCTGTG 669
Db        218 Glu------------ValPheSerSerSerThrSerAspValSerSerLeuLeuSerSer 233

Qy        670 TGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATG 729
Db        234 ThrSerSerProAlaSerSerThrIleSerGluThrLeuProPheSerSerThrIle--- 252
```



RESULT 15
PYGO_DROME
ID PYGO_DROME STANDARD; PRT; 815 AA.
AC Q9V9W8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pygopus protein (Gammy legs protein).
GN PYGO OR GAM OR CG11518.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21952490; PubMed=11955446;
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuellig S., Basler K.;
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT pygopus to the nuclear beta-catenin-TCF complex.";
RL Cell 109:47-60(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Head;
RX MEDLINE=22010046; PubMed=12015286;
RA Parker D.S., Jemison J., Cadigan K.M.;
RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling
RT in Drosophila.";
RL Development 129:2565-2576(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

```
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC   -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and
CC       thereby becomes part of the nuclear ARM/PAN complex.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Ubiquitous througout embryogenesis and larval
CC       development.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC       throughout development.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF457206; AAL91369.1; -.
DR   EMBL; AY075095; AAL79357.1; -.
DR   EMBL; AE003778; AAF57161.1; -.
DR   EMBL; AY058500; AAL13729.1; -.
DR   FlyBase; FBgn0043900; pygo.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0030528; F:transcription regulator activity; IPI.
DR   GO; GO:0030177; P:positive regulation of Wnt receptor signali. . .; IPI.
DR   GO; GO:0007367; P:segment polarity determination; IMP.
DR   GO; GO:0016055; P:Wnt receptor signaling pathway; IGI.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   SMART; SM00249; PHD; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Nuclear protein; Trans-acting factor; Wnt signaling pathway; Zinc;
KW   Metal-binding; Zinc-finger; Segmentation polarity protein.
FT   DOMAIN       39     45       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     747    805       PHD-TYPE.
FT   DOMAIN       48     65       ALA-RICH.
FT   DOMAIN      123    749       ASN/GLY/HIS/MET/PRO-RICH.
FT   CONFLICT    393    393       S -> P (IN REF. 1).
SQ   SEQUENCE   815 AA;  80493 MW;  369FD5A5D34BC136 CRC64;
```

```
Alignment Scores:
Pred. No.:              0.00652     Length:          815
Score:                  186.00      Matches:         161
Percent Similarity:     33.64%      Conservative:    59
Best Local Similarity:  24.62%      Mismatches:      222
Query Match:            4.59%       Indels:          212
DB:                     1           Gaps:            35

US-09-914-958B-35 (1-2177) x PYGO_DROME (1-815)

Qy         11 GAGGCCAGCCCGGGACCGGGGCTGGGA-----------GCAAGCAGGCGGCGGCGCCGG 58
              :::     |||||||||||||:::|||                  :::    |||      ||
Db        197 GlnMetGlyProGlyProGlyValGlyLeuProProHisMetAsnHis-GlyArgProGl 216

Qy         59 CGGCAGAGGCGGCAGCGAGCGCCCGCTTCCC------ACGCCCCTAGGCGGCGGGGCCGA 112
              ||||  ||||||         |||:::|||      :::|||:::||||||  |||
Db        216 yGlyProGlyGlyProGlyGlyProValProMetGlySerProMetGlyGlyIleAlaGl 236

Qy        113 GAGCGGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCT 172
              |||   |||:::|||            |||         :::   |||      ::
Db        236 yMetGlyGlyMetSerProMetGlyGlyMetGlyGlyProSerIleSerProHisHisMe 256

Qy        173 G---CTGCTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCAC 229
              :   :::   |||      |||          ||||||:::|||   ||||||   ::::
Db        256 tGlyMet-GlyGlyLeuSerProMetGlyGlyGlyProAsnGlyProAsnProArgAlaM 276

Qy        230 TGCAG---TCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAA 286
              ::|||   ||||||
Db        276 etGlnGlySerPro---------------------------------------------- 280

Qy        287 CTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGG 346

Db        280 ------------------------------------------------------------ 280

Qy        347 GTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACAT 406
                                          |||   ||||||                :
Db        281 ----------------------------MetGlyGlyProGlyGln-----------A 287

Qy        407 CGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCAT 466
              ::      |||::::::           :::      |||         |||:::||||
Db        287 snSerProMetAsnSerLeuProMetGlySerProMetGlyAsnProIleGlySerProL 307

Qy        467 TGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACAC 526
              |||||                  |||  ||||||              ||||||
Db        307 euGlyProProSerGlyProGlyPro-GlyAsnProGlyAsnThrGlyGlyProGln--- 325

Qy        527 CAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAA 586
                                        :::|||     |||   ||||||
Db        326 --------------------------------GlnGlnGlnGlnGlnProProGlnPro 334

Qy        587 CCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTT 646
              |||:::                  ||||||  ||||||          |||
Db        335 ProMetAsnAsn--------------GlyGlnMetGlyPro---------ProPro--- 345

Qy        647 TGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACAT 706
                                                                  ||||||
Db        346 ------------------------------------------------------LeuHis 347

Qy        707 TGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGC 766
                 |||                                    |||||||||   |||
Db        348 SerProLeuGly--------------------------------AsnGlyProThrGly 356

Qy        767 CAG-----------------TGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGC 808
                                  |||:::||||||      |||           |||
Db        357 HisGlySerHisMetProGlyGlyProIleProGly---ProGlyProGlyProGlyGly 375

Qy        809 TGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCT 868
                 |||  |||            |||                |||   :::|||
Db        376 LeuValGlyProGly--------------GlyIleSerProAlaHisGlyAsnAsnPro 390

Qy        869 CAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCA 928
                 |||   :::::::::         |||      |||   |||         ::
```

```
Db        391 -GlyGlySerGlyAsnAsnMetLeuGlyGlyAsnProGlyGlyGlyAsn------SerAs 408

Qy        929 CAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCA 988
              :|||   |||      |||         :::   ||||||            :::|||||
Db        408 nAsnAsnGlySerAsnThr---------SerAsnAlaSerAsnAsnAsnGlnAsnProHi 425

Qy        989 C---------CAGCATGGGCGA---GGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAA 1036
              |                 ||||||   |||||| |||
Db        425 sLeuSerProAlaAlaGlyArgLeuGlyVal-Pro------------------------ 436

Qy       1037 AGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCAC 1096
                   ||||||      :::   |||      |||:::   |||
Db        437 -----ThrSerMetGlnSerAsnGly------ProSerValSer---------------- 447

Qy       1097 CCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGT 1156
                        ||||||   :::   |||   |||::::::
Db        448 --------SerValAlaSerSerSerValProSerProAlaThr---------------- 459

Qy       1157 CCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAA 1216
                           ||||||||||||   |||   :::      |||   ||||||:
Db        460 ----------------ProThrLeuThrProThrSerThrAlaThrSerMetSerThrS 474

Qy       1217 CATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCG------TCATCCGTAGAGA 1270
              :::::||||||:::::|||   |||   |||   |||          |||      :
Db        474 erValProThrSerSerProAlaProProAlaMetSerProHisHisSerLeuAsnSerA 494

Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
              ::      :::|||         :::|||   |||   :::|||            ||||
Db        494 laGlyProSerProGlyMetProAsnSerGlyPro---SerProLeuGlnSerProAlaG 513

Qy       1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
              ||      |||        |||   :::      :::           ::::::   |
Db        513 lyProAsnGly---------ProAsnAsnAsnAsnSerAsn------AsnAsnAsnGlyP 528

Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAA 1450
              |||||          :::         |||::::::     ::::::
Db        528 roMetMetGlyGlnMetIle---------ProAsnAlaValProMetGlnHisGlnGlnH 545

Qy       1451 CAACTGGCTGCGAGTGGTGCCACGCACCC---------GGTTTGGGGCCTTTGCCAGGG- 1500
                 |||                 |||         |||   ||||||:::||||||
Db        545 isMetGly-----------GlyGlyProProGlyHisGlyProGlyProMetProGlyM 561

Qy       1501 ----------GAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAG--- 1546
                       ::::::                |||      ::: |||||||||
Db        561 etGlyMetAsnGlnMetLeuProProGlnGlnProSerHisLeu-GlyProProHisPro 580

Qy       1547 -------------------------GATCATCGACGGGGGCTCAGGCTACCTGTGTGA 1579
                                        |||   |||   |||
Db        581 AsnMetMetAsnHisProHisHisProHisHisHisProGlyGlyProProProHisMet 600

Qy       1580 GATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGAC 1639
                 |||      |||                |||                |||||
Db        601 MetGlyGlyProGlyMetHisGlyGlyProAlaGlyMetProProHisMetGlyGlyGly 620

Qy       1640 GTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGT--- 1696
                           |||   |||      |||      |||      :::
Db        621 ProAsnProHisMetMetGlyGlyPro---HisGlyAsnAlaGlyProHisMetGlyHis 639

Qy       1697 ---------GCTGGAGATCCTCTACCC---CCGGGATGAGGACACACTTCAGGACCCAGC 1744
                      |||   |||   |||   ||||||   |||      :::||||||
Db        640 GlyHisMetGlyGlyValProGlyProGlyProGlyProGlyGlyMetAsnGlyProPro 659

Qy       1745 CCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAA 1804
                    :::   |||                              |||||||||
Db        660 HisProHisMetSerProHis--------------------------HisGlyHisPro 670

Qy       1805 TGAATGCATCCAGTTCCCATTCGTGTGCCC 1834
                    |||      |||:::      |||
Db        671 HisHisHisHisAsnProMetGlyGlyPro 680


Search completed: August  3, 2004, 19:44:37
Job time : 83 secs
```

GenCore version 5.1.6

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 3, 2004, 19:37:45 ; Search time 109 Seconds
(without alignments)
12603.360 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 4056
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09914958/runat_03082004_161651_29683/app_query.fasta_1.2375
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914958_@CGN_1_1_122_@runat_03082004_161651_29683 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3233.5	79.7	661	4	Q9NQ80	Q9nq80 homo sapien
	2	3190.5	78.7	653	4	Q9NQ79	Q9nq79 homo sapien
	3	3181.5	78.4	637	4	Q8TE52	Q8te52 homo sapien
	4	3176.5	78.3	650	4	Q9NQ78	Q9nq78 homo sapien
	5	3046	75.1	646	11	Q8R555	Q8r555 mus musculu
	6	3035	74.8	646	11	Q8BMF1	Q8bmf1 mus musculu
	7	2133.5	52.6	451	4	Q8N4H6	Q8n4h6 homo sapien
	8	2035.5	50.2	435	11	Q8R3V8	Q8r3v8 mus musculu
	9	1967.5	48.5	418	4	Q9NW46	Q9nw46 homo sapien
	10	910	22.4	193	11	Q9QX63	Q9qx63 rattus norv
	11	265	6.5	1225	5	Q9VR49	Q9vr49 drosophila
	12	264	6.5	1081	16	Q7ULU0	Q7ulu0 rhodopirell
	13	260.5	6.4	1028	16	Q7UPI5	Q7upi5 rhodopirell
	14	255.5	6.3	800	3	Q8TFG4	Q8tfg4 schizosacch
	15	249	6.1	528	6	Q29071	Q29071 sus scrofa
	16	249	6.1	1032	16	Q7UUC1	Q7uuc1 rhodopirell
	17	248.5	6.1	615	16	Q7UJS7	Q7ujs7 rhodopirell
	18	242	6.0	2232	5	Q8IFX6	Q8ifx6 caenorhabdi
c	19	240	5.8	1637	6	Q9XSV8	Q9xsv8 bos taurus
c	20	240	5.8	5146	6	Q8SPM4	Q8spm4 bos taurus
	21	238	5.9	1806	5	Q869R4	Q869r4 dictyosteli
	22	234	5.8	1052	16	Q7UUB8	Q7uub8 rhodopirell
	23	234	5.8	1349	4	Q8WWQ4	Q8wwq4 homo sapien
	24	231.5	5.7	1237	2	Q7WYN2	Q7wyn2 acetivibrio
	25	226	5.6	1325	5	Q9BKV7	Q9bkv7 leishmania
	26	225	5.5	1107	4	Q9H4D6	Q9h4d6 homo sapien
	27	224.5	5.5	1157	4	Q9GZM2	Q9gzm2 homo sapien
	28	224.5	5.5	1199	4	Q9H487	Q9h487 homo sapien
	29	224.5	5.5	1201	4	Q9H486	Q9h486 homo sapien
	30	224.5	5.5	1214	4	Q9H485	Q9h485 homo sapien
	31	224.5	5.5	1215	4	Q9H484	Q9h484 homo sapien
	32	224.5	5.5	1256	4	Q9H483	Q9h483 homo sapien
	33	224.5	5.5	1827	4	Q9H482	Q9h482 homo sapien
	34	224.5	5.5	2117	4	Q9H4D8	Q9h4d8 homo sapien
	35	224.5	5.5	2167	4	Q9H481	Q9h481 homo sapien
	36	224.5	5.5	2169	4	Q9NY09	Q9ny09 homo sapien
	37	222.5	5.5	1187	4	Q9GZV6	Q9gzv6 homo sapien
	38	222	5.5	769	5	Q17921	Q17921 caenorhabdi
	39	219.5	5.4	1037	16	Q7UU36	Q7uu36 rhodopirell
	40	219	5.4	716	4	Q9NYE4	Q9nye4 homo sapien
	41	218.5	5.4	944	4	Q9UP82	Q9up82 homo sapien
	42	217.5	5.4	1045	4	O76037	O76037 homo sapien
	43	215.5	5.3	825	5	Q7YZW4	Q7yzw4 caenorhabdi
	44	214.5	5.3	814	11	Q80UR5	Q80ur5 mus musculu
	45	214.5	5.3	1032	5	P91365	P91365 caenorhabdi

ALIGNMENTS

RESULT 1
Q9NQ80
ID Q9NQ80 PRELIMINARY; PRT; 661 AA.
AC Q9NQ80;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ASPIC precursor.
GN ASPIC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RA Bolton M.C., Wait R., Saklatvala J.;
RT "Cloning of ASPIC, a novel protein secreted by human normal and
RT osteoarthritic cartilage, identified by 2D electrophoresis and mass
RT spectrometry.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ276171; CAB98267.1; -.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.

```
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   SIGNAL       12     27       POTENTIAL.
FT   SIGNAL        9     27       POTENTIAL.
FT   CHAIN        28    661       ASPIC.
SQ   SEQUENCE   661 AA;  71420 MW;  2EF75321817F0D01 CRC64;

Alignment Scores:
Pred. No.:                2.35e-193      Length:          661
Score:                    3233.50        Matches:         611
Percent Similarity:       97.92%         Conservative:    2
Best Local Similarity:    97.60%         Mismatches:      2
Query Match:              79.72%         Indels:          11
DB:                       4              Gaps:            1

US-09-914-958B-35 (1-2177) x Q9NQ80 (1-661)

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAlaProSerAlaAspProGlyMetSerArgMetLeuProPheLeuLeuLeuLeuTrp 20

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheLeuProIleThrGluGlySerGlnArgAlaGluProMetPheThrAlaValThrAsn 40

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerValLeuProProAspTyrAspSerAsnProThrGlnLeuAsnTyrGlyValAlaVal 60

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ThrAspValAspHisAspGlyAspPheGluIleValValAlaGlyTyrAsnGlyProAsn 80

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuValLeuLysTyrAspArgAlaGlnLysArgLeuValAsnIleAlaValAspGluArg 100

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 SerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAlaCys 120

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAsnAlaPheSer 140

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 GlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAspIle 160

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuSerAspGluValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSerVal 180

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 722
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AlaCysValAspArgLysGlySerGlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyr 200

Qy        723 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGG 782
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GlyAsnValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSerArg 220

Qy        783 GGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGC 842
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrGlyGly 240

Qy        843 CGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAAT 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgGlyValSerValGlyProIleLeuSerSerSerAlaSerAspIlePheCysAspAsn 260

Qy        903 GAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCG 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GluAsnGlyProAsnPheLeuPheHisAsnArgGlyAspGlyThrPheValAspAlaAla 280

Qy        963 GCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTC 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAspPhe 300

Qy       1023 AACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTAT 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeuTyr 320

Qy       1083 CTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCC 1142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 LeuGlnMetSerThrHisGlyLysValArgPheArgAspIleAlaSerProLysPheSer 340

Qy       1143 ATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAG 1202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 MetProSerProValArgThrValIleThrAlaAspPheAspAsnAspGlnGluLeuGlu 360

Qy       1203 ATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATC 1262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 IlePhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgValIle 380

Qy       1263 CGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCT 1322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ArgArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGluPro 400

Qy       1323 GAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTC 1382
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GluGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAspLeu 420

Qy       1383 ATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGC 1442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGlnGly 440

Qy       1443 TTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGA 1502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 PheAsnAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArgGly 460

Qy       1503 GCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGC 1562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGlyGly 480

Qy       1563 TCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCC 1622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyLysAspGluAla 500

Qy       1623 AGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGG 1682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 SerSerValGluValThrTrpProAspGlyLysMetValSerArgAsnValAlaSerGly 520

Qy       1683 GAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCA 1742
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 GluMetAsnSerValLeuGluIleLeuTyrProArgAspGluAspThrLeuGlnAspPro 540

Qy       1743 GCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACC 1802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AlaProLeuGluCysGlyGlnGlyPheSerGlnGlnGluAsnGlyHisCysMetAspThr 560

Qy       1803 AATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACC 1862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsnThr 580

Qy       1863 TATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAG 1922
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 TyrGlySerTyrArgCysArgThrAsnLysLysCysSerArgGlyTyrGluProAsnGlu 600

Qy       1923 GATGGCACAGCCTGCGTGGGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTT 1982
              ||||||||||||||||||                                      :::
Db        601 AspGlyThrAlaCysValGly----------------------------------ThrLeu 609
```

```
Qy       1983 GGGAAGAGCCTTGGTCCC 2000
              |||:::|||   ||||||
Db        610 GlyGlnSerProGlyPro 615

RESULT 2
Q9NQ79
ID   Q9NQ79      PRELIMINARY;      PRT;   653 AA.
AC   Q9NQ79;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ASPIC precursor (Chondrocyte expressed protein 68 kDa).
GN   ASPIC1 OR CEP-68.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RA   Bolton M.C., Wait R., Saklatvala J.;
RT   "Cloning of ASPIC, a novel protein secreted by human normal and
RT   osteoarthritic cartilage, identified by 2D electrophoresis and mass
RT   spectrometry.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RA   Steck E., Gress T., Loew M., Richter W.;
RT   "CEP-68: a novel human marker gene specifically expressed in cultured
RT   chondrocytes.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ276171; CAB98268.1; -.
DR   EMBL; AJ279016; CAC08451.1; -.
DR   Genew; HGNC:14882; CRTAC1.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    653       ASPIC.
SQ   SEQUENCE   653 AA;  70693 MW;  3634B98E8ACE3CBA CRC64;

Alignment Scores:
Pred. No.:              1.12e-190      Length:          653
Score:                  3190.50        Matches:         603
Percent Similarity:     97.90%         Conservative:    2
Best Local Similarity:  97.57%         Mismatches:      2
Query Match:            78.66%         Indels:          11
DB:                     4              Gaps:            1

US-09-914-958B-35 (1-2177) x Q9NQ79 (1-653)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440
```

```
Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGG 1946
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--- 599

Qy       1947 TGGAGCCCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
Db        600 ----------------------------ThrLeuGlyGlnSerProGlyPro 607
```

RESULT 3
Q8TE52
ID Q8TE52 PRELIMINARY; PRT; 637 AA.
AC Q8TE52;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CRTAC1-B protein.
GN CRTAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Steck E., Richter W.;
RT "A novel form of CRTAC1 expressed in human and mouse brain generated
RT by alternative splicing of a newly identified last exon.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ421515; CAD13394.1; -.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR PROSITE; PS01187; EGF_CA; 1.
SQ SEQUENCE 637 AA; 69706 MW; 42543412E0419D6E CRC64;

Alignment Scores:

Pred. No.:	4.06e-190	Length:	637
Score:	3181.50	Matches:	602
Percent Similarity:	97.89%	Conservative:	2
Best Local Similarity:	97.57%	Mismatches:	6
Query Match:	78.44%	Indels:	7

```
DB:                    4              Gaps:          1

US-09-914-958B-35 (1-2177) x Q8TE52 (1-637)

Qy        147 ATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCC 206
Db          1 MetSerArgMetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySer 20

Qy        207 CAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGAC 266
Db         21 GlnArgAlaGluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAsp 40

Qy        267 AGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGAC 326
Db         41 SerAsnProThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAsp 60

Qy        327 TTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCC 386
Db         61 PheGluIleValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAla 80

Qy        387 CAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGG 446
Db         81 GlnLysArgLeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArg 100

Qy        447 GACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAG 506
Db        101 AspArgGlnGlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGlu 120

Qy        507 GAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAG 566
Db        121 GluIleTyrPheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLys 140

Qy        567 TTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCC 626
Db        141 LeuPheLysPheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAla 160

Qy        627 CGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCT 686
Db        161 ArgGlyValAlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySer 180

Qy        687 GGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTC 746
Db        181 GlyArgTyrSerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeu 200

Qy        747 ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTG 806
Db        201 IleGluMetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspVal 220

Qy        807 GCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATC 866
Db        221 AlaAlaGluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIle 240

Qy        867 CTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTC 926
Db        241 LeuSerSerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPhe 260

Qy        927 CACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCC 986
Db        261 HisAsnArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspPro 280

Qy        987 CACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATC 1046
Db        281 HisGlnHisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIle 300

Qy       1047 GTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAG 1106
Db        301 ValTyrGlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLys 320

Qy       1107 GTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTC 1166
Db        321 ValArgPheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrVal 340

Qy       1167 ATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTAC 1226
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleThrAlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyr 360

Qy       1227 CGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTC 1286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ArgSerSerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeu 380

Qy       1287 ATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTG 1346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleGluGluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyVal 400

Qy       1347 GTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATG 1406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ValThrAspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMet 420

Qy       1407 GCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG 1466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AlaGlnProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgVal 440

Qy       1467 GTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAG 1526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLys 460

Qy       1527 AAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAG 1586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 LysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGlu 480

Qy       1587 CCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCA 1646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ProValAlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpPro 500

Qy       1647 GATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATC 1706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AspGlyLysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIle 520

Qy       1707 CTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGA 1766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 LeuTyrProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGly 540

Qy       1767 TTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTC 1826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PheSerGlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPhe 560

Qy       1827 GTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACC 1886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThr 580

Qy       1887 AACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG------ 1940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 AsnLysLysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGln 600

Qy       1941 ---------------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976
                             |||:::   |||        :::|||       |||
Db        601 ValAlaPheLeuGlyGlyTyrSerSerAlaAlaSerArgIleSerGluPro 617
```

```
RESULT 4
Q9NQ78
ID   Q9NQ78      PRELIMINARY;      PRT;   650 AA.
AC   Q9NQ78;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ASPIC precursor.
GN   ASPIC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RA   Bolton M.C., Wait R., Saklatvala J.;
RT   "Cloning of ASPIC, a novel protein secreted by human normal and
RT   osteoarthritic cartilage, identified by 2D electrophoresis and mass
RT   spectrometry.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ276171; CAB98269.1; -.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    650       ASPIC.
SQ   SEQUENCE   650 AA;  70319 MW;  9EF452AD76CAC39F CRC64;

Alignment Scores:
Pred. No.:              8.38e-190     Length:          650
Score:                  3176.50       Matches:         600
Percent Similarity:     97.89%        Conservative:    2
Best Local Similarity:  97.56%        Mismatches:      2
Query Match:            78.32%        Indels:          11
DB:                     4             Gaps:            1

US-09-914-958B-35 (1-2177) x Q9NQ78 (1-650)
```

```
Qy        156 ATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCT 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLeuProPheLeuLeuLeuLeuTrpPheLeuProIleThrGluGlySerGlnArgAla 20

Qy        216 GAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCC 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluProMetPheThrAlaValThrAsnSerValLeuProProAspTyrAspSerAsnPro 40

Qy        276 ACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATC 335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ThrGlnLeuAsnTyrGlyValAlaValThrAspValAspHisAspGlyAspPheGluIle 60

Qy        336 GTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGG 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ValValAlaGlyTyrAsnGlyProAsnLeuValLeuLysTyrAspArgAlaGlnLysArg 80

Qy        396 CTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAG 455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LeuValAsnIleAlaValAspGluArgSerSerProTyrTyrAlaLeuArgAspArgGln 100

Qy        456 GGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTAC 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnAlaIleGlyValThrAlaCysAspIleAspGlyAspGlyArgGluGluIleTyr 120

Qy        516 TTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAG 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheLeuAsnThrAsnAsnAlaPheSerGlyValAlaThrTyrThrAspLysLeuPheLys 140

Qy        576 TTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTG 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 PheArgAsnAsnArgTrpGluAspIleLeuSerAspGluValAsnValAlaArgGlyVal 160

Qy        636 GCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTAC 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 AlaSerLeuPheAlaGlyArgSerValAlaCysValAspArgLysGlySerGlyArgTyr 180

Qy        696 TCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATG 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SerIleTyrIleAlaAsnTyrAlaTyrGlyAsnValGlyProAspAlaLeuIleGluMet 200

Qy        756 GACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAG 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAlaGlu 220

Qy        816 GCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGC 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        221 AlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSerSer 240

Qy        876 AGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGG 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsnArg 260

Qy        936 GGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCAT 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGlnHis 280

Qy        996 GGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGC 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyrGly 300

Qy       1056 AACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTC 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArgPhe 320

Qy       1116 CGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCC 1175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThrAla 340

Qy       1176 GACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCC 1235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 AspPheAspAsnAspGlnGluLeuGluIlePheAsnAsnIleAlaTyrArgSerSer 360

Qy       1236 TCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAG 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGluGlu 380

Qy       1296 CTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGAC 1355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThrAsp 400

Qy       1356 TTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCG 1415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 PheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGlnPro 420

Qy       1416 CTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGC 1475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValProArg 440

Qy       1476 ACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGG 1535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 ThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSerGly 460

Qy       1536 GCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCA 1595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 AlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProValAla 480

Qy       1596 CACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAG 1655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGlyLys 500

Qy       1656 ATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCC 1715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 MetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyrPro 520

Qy       1716 CGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAG 1775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 ArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSerGln 540

Qy       1776 CAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCT 1835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCysPro 560

Qy       1836 CGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAG 1895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLysLys 580

Qy       1896 TGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCCCT 1955
              ||||||||||||||||||||||||||||||||||||||||||||
Db        581 CysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly---------- 596
```

```
Qy       1956 GTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                        :::|||:::|||   ||||||
Db        597 --------------------ThrLeuGlyGlnSerProGlyPro 604

RESULT 5
Q8R555
ID   Q8R555      PRELIMINARY;      PRT;   646 AA.
AC   Q8R555;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CRTAC1-B protein (ASPIC precursor homolog).
GN   2810454P21RIK OR CRTAC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Steck E., Richter W.;
RT   "A novel form of CRTAC1 expressed in human and mouse brain generated
RT   by alternative splicing of a newly identified last exon.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Spinal cord;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AJ421516; CAD13395.1; -.
DR   EMBL; AK049801; BAC33924.1; -.
DR   MGD; MGI:1920082; Crtac1.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   646 AA;  70324 MW;  154C45E636DF2D98 CRC64;

Alignment Scores:
Pred. No.:              1.16e-181      Length:          646
Score:                  3046.00        Matches:         576
Percent Similarity:     94.89%         Conservative:    18
Best Local Similarity:  92.01%         Mismatches:      24
Query Match:            75.10%         Indels:          8
DB:                     11             Gaps:            2

US-09-914-958B-35 (1-2177) x Q8R555 (1-646)

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTG---CTGCTC 179
              ||||||||||||||||||||||||||||   ||||||      ||||||      |||
Db          1 MetAlaProSerAlaAspProGlyMetValArgMetAlaLeuLeuLeuLeuProProLeu 20

Qy        180 TGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACC 239
              |||   ||||||:::|||   ||||||||||||||||||||||||||||||||||||||
Db         21 TrpLeuLeuProLeuThrGlyGlySerGlnArgAlaGluProMetPheThrAlaValThr 40

Qy        240 AACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCA 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AsnSerValLeuProProAspTyrAspSerAsnProThrGlnLeuAsnTyrGlyValAla 60

Qy        300 GTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCC 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ValThrAspValAspHisAspGlyAspPheGluIleValValAlaGlyTyrThrGlyPro 80
```

```
Qy       360 AACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAG 419
Db        81 AsnLeuValLeuLysTyrAsnArgAlaGlnAsnArgLeuValAsnIleAlaValAspGlu 100

Qy       420 CGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCC 479
Db       101 ArgSerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAla 120

Qy       480 TGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTC 539
Db       121 CysAspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAsnAlaPhe 140

Qy       540 TCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGAC 599
Db       141 SerGlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAsp 160

Qy       600 ATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCT 659
Db       161 IleLeuSerAspAspValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSer 180

Qy       660 GTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCC 719
Db       181 ValAlaCysValAspArgThrGlySerGlyArgTyrSerIleTyrIleAlaAsnTyrAla 200

Qy       720 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 779
Db       201 TyrGlyAspValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSer 220

Qy       780 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 839
Db       221 ArgGlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrAla 240

Qy       840 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 899
Db       241 GlyArgGlyValSerValGlyProIleLeuSerSerSerAlaSerAspIlePheCysAsp 260

Qy       900 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 959
Db       261 AsnGluAsnGlyProAsnPheLeuPheHisAsnGlnGlyAsnGlyThrPheValAspThr 280

Qy       960 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 1019
Db       281 AlaAlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAsp 300

Qy      1020 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 1079
Db       301 PheAsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeu 320

Qy      1080 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 1139
Db       321 TyrLeuGlnMetSerAlaHisGlyLysValArgPheArgAspIleAlaSerProLysPhe 340

Qy      1140 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 1199
Db       341 SerThrProSerProValArgThrValIleAlaAlaAspPheAspAsnAspGlnGluLeu 360

Qy      1200 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 1259
Db       361 GluValPhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgVal 380

Qy      1260 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 1319
Db       381 IleArgArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGlu 400

Qy      1320 CCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGAC 1379
Db       401 ProGluGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAsp 420

Qy      1380 CTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAG 1439
Db       421 LeuIleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGln 440

Qy      1440 GGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGG 1499
Db       441 GlyPheSerAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArg 460

Qy      1500 GGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGG 1559
Db       461 GlyAlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGly 480

Qy      1560 GGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAA 1619
Db       481 GlySerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyArgAspGlu 500

Qy      1620 GCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGC 1679
Db       501 AlaSerSerValGluValThrTrpProAspGlyLysMetValSerArgSerValAlaSer 520

Qy      1680 GGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGAC 1739
Db       521 GluGluMetAsnSerValLeuGluIleLeuTyrProGlnAspGluAspLysLeuGlnAsn 540

Qy      1740 CCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGAC 1799
Db       541 ThrAlaProLeuGluCysGlyGlnGlyPheSerGlnGlnAspAsnGlyHisCysMetAsp 560

Qy      1800 ACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAAC 1859
Db       561 ThrAsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsn 580

Qy      1860 ACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAAC 1919
Db       581 ThrTyrGlySerTyrArgCysArgThrAsnLysArgCysAsnArgGlyTyrGluProAsn 600

Qy      1920 GAGGATGGCACAGCCTGCGTG----------GGCTGGTGGAGCCCTGTG 1958
Db       601 GluAspGlyThrAlaCysValAlaGlnValAlaPheLeuGlyGlyTyrSerSerAlaAla 620

Qy      1959 TTGAAGATAGTGACACCA 1976
Db       621 PheArgLeuSerGluPro 626
```

RESULT 6
Q8BMF1
ID   Q8BMF1   PRELIMINARY;   PRT;   646 AA.
AC   Q8BMF1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ASPIC precursor homolog.
GN   CRTAC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK032328; BAC27817.1; -.
DR   MGD; MGI:1920082; Crtac1.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   646 AA; 70197 MW; B8841A37B05FF314 CRC64;

```
Alignment Scores:
Pred. No.:                5.61e-181      Length:          646
Score:                    3035.00        Matches:         574
Percent Similarity:       94.73%         Conservative:    19
Best Local Similarity:    91.69%         Mismatches:      25
Query Match:              74.83%         Indels:          8
DB:                       11             Gaps:            2

US-09-914-958B-35 (1-2177) x Q8BMF1 (1-646)

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTG---CTGCTC 179
              ||||||||||||||||||||||||||||||    ||||||          ||||||      |||
Db          1 MetAlaProSerAlaAspProGlyMetValArgMetAlaLeuLeuLeuLeuProProLeu 20

Qy        180 TGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACC 239
              |||   ||||||:::|||   ||||||||||||||||||||||||||||||||||||||||
Db         21 TrpLeuLeuProLeuThrGlySerGlnArgAlaGluProMetPheThrAlaValThr 40

Qy        240 AACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCA 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 AsnSerValLeuProProAspTyrAspSerAsnProThrGlnLeuAsnTyrGlyValAla 60

Qy        300 GTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCC 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db         61 ValThrAspValAspHisAspGlyAspPheGluIleValValAlaGlyTyrThrGlyPro 80

Qy        360 AACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAG 419
              ||||||||||||||||:::|||||||||   ||||||||||||||||||||||||||||||
Db         81 AsnLeuValLeuLysTyrAsnArgAlaGlnAsnArgLeuValAsnIleAlaValAspGlu 100

Qy        420 CGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 ArgSerSerProTyrTyrAlaLeuArgAspArgGlnGlyAsnAlaIleGlyValThrAla 120

Qy        480 TGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTC 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CysAspIleAspGlyAspGlyArgGluGluIleTyrPheLeuAsnThrAsnAsnAlaPhe 140

Qy        540 TCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGAC 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 SerGlyValAlaThrTyrThrAspLysLeuPheLysPheArgAsnAsnArgTrpGluAsp 160

Qy        600 ATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCT 659
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        161 IleLeuSerAspAspValAsnValAlaArgGlyValAlaSerLeuPheAlaGlyArgSer 180

Qy        660 GTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCC 719
              ||||||||||||||||    |||||||||   ||||||||||||||||||||||||||||
Db        181 ValAlaCysValAspArgThrGlySerGlyGlyTyrSerIleTyrIleAlaAsnTyrAla 200

Qy        720 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 779
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TyrGlyAspValGlyProAspAlaLeuIleGluMetAspProGluAlaSerAspLeuSer 220

Qy        780 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ArgGlyIleLeuAlaLeuArgAspValAlaAlaGluAlaGlyValSerLysTyrThrAla 240

Qy        840 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GlyArgGlyValSerValGlyProIleLeuSerSerSerAlaSerAspIlePheCysAsp 260

Qy        900 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 959
              ||||||||||||||||||||||||||||||||:::|||:::|||||||||||||||||||
Db        261 AsnGluAsnGlyProAsnPheLeuPheHisAsnGlnGlyAsnGlyThrPheValAspThr 280

Qy        960 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaAlaSerAlaGlyValAspAspProHisGlnHisGlyArgGlyValAlaLeuAlaAsp 300

Qy       1020 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        301 PheAsnArgAspGlyLysValAspIleValTyrGlyAsnTrpAsnGlyProHisArgLeu 320

Qy       1080 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 1139
              ||||||||||||||    ||||||||||||:::||||||||||||||||||||||||||||
Db        321 TyrLeuGlnMetSerAlaHisGlyLysValGlnPheArgAspIleAlaSerProLysPhe 340

Qy       1140 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 1199
              |||   ||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        341 SerThrProSerProValArgThrValIleAlaAlaAspPheAspAsnAspGlnGluLeu 360

Qy       1200 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 1259
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluValPhePheAsnAsnIleAlaTyrArgSerSerSerAlaAsnArgLeuPheArgVal 380

Qy       1260 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 IleArgArgGluHisGlyAspProLeuIleGluGluLeuAsnProGlyAspAlaLeuGlu 400

Qy       1320 CCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGAC 1379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 ProGluGlyArgGlyThrGlyGlyValValThrAspPheAspGlyAspGlyMetLeuAsp 420

Qy       1380 CTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAG 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LeuIleLeuSerHisGlyGluSerMetAlaGlnProLeuSerValPheArgGlyAsnGln 440

Qy       1440 GGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGG 1499
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 GlyPheSerAsnAsnTrpLeuArgValValProArgThrArgPheGlyAlaPheAlaArg 460

Qy       1500 GGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGG 1559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 GlyAlaLysValValLeuTyrThrLysLysSerGlyAlaHisLeuArgIleIleAspGly 480

Qy       1560 GGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAA 1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        481 GlySerGlyTyrLeuCysGluMetGluProValAlaHisPheGlyLeuGlyArgAspGlu 500

Qy       1620 GCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGC 1679
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        501 AlaSerSerValGluValThrTrpProAspGlyLysMetValSerArgSerValAlaSer 520

Qy       1680 GGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGAC 1739
              ||||||||||||||||||||||||||||||:::|||||||||   |||||||:::
Db        521 GluGluMetAsnSerValLeuGluIleLeuTyrProGlnAspGluAspLysLeuGlnAsn 540

Qy       1740 CCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGAC 1799
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        541 ThrAlaProLeuGluCysGlyGlnGlyPheSerGlnGlnAspAsnGlyHisCysMetAsp 560

Qy       1800 ACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAAC 1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 ThrAsnGluCysIleGlnPheProPheValCysProArgAspLysProValCysValAsn 580

Qy       1860 ACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAAC 1919
              ||||||||||||||||||||||||||||||||:::|||:::|||||||||||||||||||
Db        581 ThrTyrGlySerTyrArgCysArgThrAsnLysArgCysAsnArgGlyTyrGluProAsn 600

Qy       1920 GAGGATGGCACAGCCTGCGTG----------------------GGCTGGTGGAGCCCTGTG 1958
              |||||||||||||||||||                      ||:::   |||
Db        601 GluAspGlyThrAlaCysValAlaGlnValAlaPheLeuGlyGlyTyrSerSerAlaAla 620

Qy       1959 TTGAAGATAGTGACACCA 1976
                  ::::::     |||
Db        621 PheArgLeuSerGluPro 626

RESULT 7
Q8N4H6
ID   Q8N4H6      PRELIMINARY;      PRT;   451 AA.
AC   Q8N4H6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```
DE   Similar to cartilage acidic protein 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034245; AAH34245.1; -.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 3.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   451 AA;  48270 MW;  26F3DE1E8D970D2F CRC64;

Alignment Scores:
Pred. No.:                8.5e-125      Length:          451
Score:                    2133.50       Matches:         401
Percent Similarity:       96.88%        Conservative:    2
Best Local Similarity:    96.39%        Mismatches:      2
Query Match:              52.60%        Indels:          11
DB:                       4             Gaps:            1

US-09-914-958B-35 (1-2177) x Q8N4H6 (1-451)

Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 ProArgAspGluAspThrLeuGlnAspProAlaProLeuGluCysGlyGlnGlyPheSer 340

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnGlnGluAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCys 360

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 LysCysSerArgGlyTyrGluProAsnGluAspGlyThrAlaCysValGly--------- 397

Qy       1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCC 2000
                                       :::|||:::|||   ||||||
Db        398 -----------------------ThrLeuGlyGlnSerProGlyPro 405

RESULT 8
Q8R3V8
ID   Q8R3V8      PRELIMINARY;      PRT;   435 AA.
AC   Q8R3V8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cartilage acidic protein 1.
GN   CRTAC1 OR 2810454P21RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024472; AAH24472.1; -.
DR   MGD; MGI:1920082; Crtac1.
DR   GO; GO:0008305; C:integrin complex; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR   GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 3.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   435 AA;  47315 MW;  01A18E8556A2F22B CRC64;

Alignment Scores:
```

Pred. No.:	1.09e-118	Length:	435
Score:	2035.50	Matches:	380
Percent Similarity:	94.94%	Conservative:	14
Best Local Similarity:	91.57%	Mismatches:	14
Query Match:	50.18%	Indels:	7
DB:	11	Gaps:	1

US-09-914-958B-35 (1-2177) x Q8R3V8 (1-435)

```
Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrAlaGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60

Qy        933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
              :::|||:::|||||||||||||||   ||||||||||||||||||||||||||||||||
Db         61 GlnGlyAsnGlyThrPheValAspThrAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy        993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy       1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerAlaHisGlyLysValArg 120

Qy       1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        121 PheArgAspIleAlaSerProLysPheSerThrProSerProValArgThrValIleAla 140

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        141 AlaAspPheAspAsnAspGlnGluLeuGluValPhePheAsnAsnIleAlaTyrArgSer 160

Qy       1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy       1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy       1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy       1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db        221 ProLeuSerValPheArgGlyAsnGlnGlyPheSerAsnAsnTrpLeuArgValValPro 240

Qy       1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy       1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy       1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        281 AlaHisPheGlyLeuGlyArgAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy       1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
              ||||||||||||||:::|||||||||   ||||||||||||||||||||||||||||||||
Db        301 LysMetValSerArgSerValAlaSerGluGluMetAsnSerValLeuGluIleLeuTyr 320

Qy       1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
              |||:::|||||||||   ||||||:::   ||||||||||||||||||||||||||||||
Db        321 ProGlnAspGluAspLysLeuGlnAsnThrAlaProLeuGluCysGlyGlnGlyPheSer 340

Qy       1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 GlnGlnAspAsnGlyHisCysMetAspThrAsnGluCysIleGlnPheProPheValCys 360

Qy       1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArgThrAsnLys 380

Qy       1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTG---------- 1940
              :::|||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        381 ArgCysAsnArgGlyTyrGluProAsnGluAspGlyThrAlaCysValAlaGlnValAla 400

Qy       1941 ---------GGCTGGTGGAGCCCTGTGTTGAAGATAGTGACACCA 1976
                      |||:::   |||         ::::::      |||
Db        401 PheLeuGlyGlyTyrSerSerAlaAlaPheArgLeuSerGluPro 415
```

RESULT 9
Q9NW46
ID Q9NW46 PRELIMINARY; PRT; 418 AA.
AC Q9NW46;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ10320.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK001182; BAA91540.1; -.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 3.
KW Hypothetical protein.
SQ SEQUENCE 418 AA; 44861 MW; B85F301A96DD545F CRC64;

Pred. No.:	1.87e-114	Length:	418
Score:	1967.50	Matches:	385
Percent Similarity:	89.81%	Conservative:	3
Best Local Similarity:	89.12%	Mismatches:	10
Query Match:	48.51%	Indels:	34
DB:	4	Gaps:	2

US-09-914-958B-35 (1-2177) x Q9NW46 (1-418)

```
Qy        753 ATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCT 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetAspProGluAlaSerAspLeuSerArgGlyIleLeuAlaLeuArgAspValAlaAla 20

Qy        813 GAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GluAlaGlyValSerLysTyrThrGlyGlyArgGlyValSerValGlyProIleLeuSer 40

Qy        873 AGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAAC 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 SerSerAlaSerAspIlePheCysAspAsnGluAsnGlyProAsnPheLeuPheHisAsn 60
```

```
Qy       933 CGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAG 992
Db        61 ArgGlyAspGlyThrPheValAspAlaAlaAlaSerAlaGlyValAspAspProHisGln 80

Qy       993 CATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTAT 1052
Db        81 HisGlyArgGlyValAlaLeuAlaAspPheAsnArgAspGlyLysValAspIleValTyr 100

Qy      1053 GGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGC 1112
Db       101 GlyAsnTrpAsnGlyProHisArgLeuTyrLeuGlnMetSerThrHisGlyLysValArg 120

Qy      1113 TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACC 1172
Db       121 PheArgAspIleAlaSerProLysPheSerMetProSerProValArgThrValIleThr 140

Qy      1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGC 1232
Db       141 AlaAspPheAspAsnAspGlnGluLeuGluIlePhePheAsnAsnIleAlaTyrArgSer 160

Qy      1233 TCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAG 1292
Db       161 SerSerAlaAsnArgLeuPheArgValIleArgArgGluHisGlyAspProLeuIleGlu 180

Qy      1293 GAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACC 1352
Db       181 GluLeuAsnProGlyAspAlaLeuGluProGluGlyArgGlyThrGlyGlyValValThr 200

Qy      1353 GACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAG 1412
Db       201 AspPheAspGlyAspGlyMetLeuAspLeuIleLeuSerHisGlyGluSerMetAlaGln 220

Qy      1413 CCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCA 1472
Db       221 ProLeuSerValPheArgGlyAsnGlnGlyPheAsnAsnAsnTrpLeuArgValValPro 240

Qy      1473 CGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGT 1532
Db       241 ArgThrArgValGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThrLysLysSer 260

Qy      1533 GGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTG 1592
Db       261 GlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMetGluProVal 280

Qy      1593 GCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC 1652
Db       281 AlaHisPheGlyLeuGlyLysAspGluAlaSerSerValGluValThrTrpProAspGly 300

Qy      1653 AAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTAC 1712
Db       301 LysMetValSerArgAsnValAlaSerGlyGluMetAsnSerValLeuGluIleLeuTyr 320

Qy      1713 CCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCC 1772
Db       321 ProArgAspGluAspThrLeuGlnAspProAlaProLeu-Glu---------------- 334

Qy      1773 CAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGC 1832
Db       335 ------------------------ThrProMetAsnAlaSerSerSerHisSerCysAl 346

Qy      1833 CCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAG 1892
Db       346 aLeuGluThrSerProTyrValSerThrProMetGluAlaThrGlyAlaGlyProThrAr 366

Qy      1893 AAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGC 1952
Db       366 gSerAlaValGlyAlaThrSerProThrArgMetAlaGlnProAlaTrpGly-------- 383

Qy      1953 CCTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAA 2012
Db       384 ----------------------------------------------------LeuSerAlaSe 387

Qy      2013 TCACTGCCTTGAATCACCGCCTGGAATACCTGTT 2046
Db       387 rHisArgAlaProAlaProProProProProLeu 398

RESULT 10
Q9QX63
ID   Q9QX63      PRELIMINARY;      PRT;   193 AA.
AC   Q9QX63;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   W307.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Kawasaki H., Housman D.E., Graybiel A.M.;
RT   "Characterization of region-specific genes from rat brain.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U78304; AAF21221.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   193 AA;  21572 MW;  B94898BFE5EE045A CRC64;

Alignment Scores:
Pred. No.:              1.32e-48      Length:          193
Score:                  910.00        Matches:         165
Percent Similarity:     96.67%        Conservative:    9
Best Local Similarity:  91.67%        Mismatches:      6
Query Match:            22.44%        Indels:          0
DB:                     11            Gaps:            0

US-09-914-958B-35 (1-2177) x Q9QX63 (1-193)

Qy      1404 ATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGA 1463
Db         1 MetAspGlnProLeuSerValPheArgGlyAsnGlnGlyPheSerAsnAsnTrpLeuArg 20

Qy      1464 GTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACC 1523
Db        21 ValValProArgThrArgPheGlyAlaPheAlaArgGlyAlaLysValValLeuTyrThr 40

Qy      1524 AAGAAGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATG 1583
Db        41 LysLysSerGlyAlaHisLeuArgIleIleAspGlyGlySerGlyTyrLeuCysGluMet 60

Qy      1584 GAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGG 1643
Db        61 GluProValAlaHisPheGlyLeuGlyArgAspGluAlaSerSerValGluValThrTrp 80

Qy      1644 CCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAG 1703
Db        81 ProAspGlyLysMetLeuSerArgSerValAlaAsnArgGluMetAsnSerValLeuGlu 100

Qy      1704 ATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAA 1763
Db       101 IleLeuTyrProArgAspGluAspLysLeuGlnAsnThrAlaProLeuGluCysGlyGln 120

Qy      1764 GGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCA 1823
Db       121 GlyPheSerGlnGlnAspAsnGlyHisCysMetAspThrAsnGluCysIleGlnPhePro 140

Qy      1824 TTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGG 1883
Db       141 PheValCysProArgAspLysProValCysValAsnThrTyrGlySerTyrArgCysArg 160

Qy      1884 ACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGC 1943
Db       161 ThrAsnLysArgCysSerArgGlyTyrGluProAsnGluAspGlyThrSerLeuArgGly 180
```

```
RESULT 11
Q9VR49
ID   Q9VR49      PRELIMINARY;      PRT;  1225 AA.
AC   Q9VR49;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG3047-PA.
GN   SGS1 OR CG3047.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA   Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003575; AAF50957.3; -.
SQ   SEQUENCE   1225 AA;  127735 MW;  A99AF9D4404C79F4 CRC64;

Alignment Scores:
Pred. No.:                4.33e-08      Length:          1225
Score:                    265.00        Matches:         160
Percent Similarity:       36.43%        Conservative:    75
Best Local Similarity:    24.81%        Mismatches:      280
Query Match:              6.53%         Indels:          130
DB:                       5             Gaps:            29

US-09-914-958B-35 (1-2177) x Q9VR49 (1-1225)

Qy      139 ACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACTG 198
            ||||||   |||                    |||   ||||||
Db      158 ThrProLeuCysThrAspThr---ProProCysThrCysSer------------------ 170

Qy      199 AGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCTG 258
                                            |||          |||:::     |||:::
Db      171 -----------------GluGluSerThrAlaIleProSerSerProCysIleAsp 184

Qy      259 ACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATG 318
            |||   ||||||||||||   |||                                   |||
Db      185 ThrSerThrValIleProThrSer--------------ProCysThrGlnGluThrThr 199

Qy      319 ATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCA------ACCTGGTTCTGA 372
                |||         |||   :::||||||                    |||
Db      200 ThrProThrProThrCysSerThrGlnGlyThrGlnThrThrProCysThrCysAlaGln 219

Qy      373 AGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCT 432
            :::   |||  |||||||||                :::  :::  |||  :::   |||
Db      220 ThrThrThrThrProArgSer-----------ThrThrThrThrSerThrSerArgPro 235

Qy      433 ACTACGCGCTGCGGGACCGGCAGGGGAACG--------------------CCATTGGGG 471
            ||||||          |||  |||  |||                        |||
Db      236 ThrThrThr------ThrProArgSerThrThrThrThrThrThrSerArgProThrThr 253

Qy      472 TCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCA--- 528
            :::   |||   ::::::|||   |||      ||||||   |||::::::||||||
Db      254 ThrThrProArgSerThrThrThrThrThrArgArgProThrThrThrThrProArg 273

Qy      529 --------------------ATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGT 567
                                    |||:::          ||||||:::   |||:::
Db      274 CysThrThrThrThrSerThrCysAlaProThrThrThrThrProArgSerThrThrThr 293

Qy      568 TGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGGCCC 627
              ::::::|||      |||      |||:::                 :::|||
Db      294 ThrThrThrSerArgProThr------ThrThrThrProArgCysThrThrThrThrSer 311

Qy      628 GTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTG 687
                |||                                   |||   |||:::
Db      312 ThrCysSerProThrArgThr----------------------ThrProArgSerThr 323

Qy      688 GACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCCTCA 747
                |||   ::::::   |||  |||  |||                    ||||||
Db      324 ThrThrThrSerThrSerArgProThrThrThrThrProArgCysThrThrThrProSer 343

Qy      748 TTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGG 807
                ||| ||||||  |||:::                 ||||||
Db      344 ThrThrCysThrArgArgProThrThrThrThrPro---------ArgSerThrThrThr 360

Qy      808 CTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCC 867
                  |||                                 :::                |||
Db      361 ThrCysThrSerGlyProThrThrThrThrProArgSerThrThrThrThrCysThrSer 380

Qy      868 ------TCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACT--- 918
                  :::          ||||||   ::::::   |||          |||   |||
Db      381 GlyProThrThrThrThrProArgSerThrThrThrThrSerThrSerGlyProThrThr 400

Qy      919 ------------TCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCA 966
```

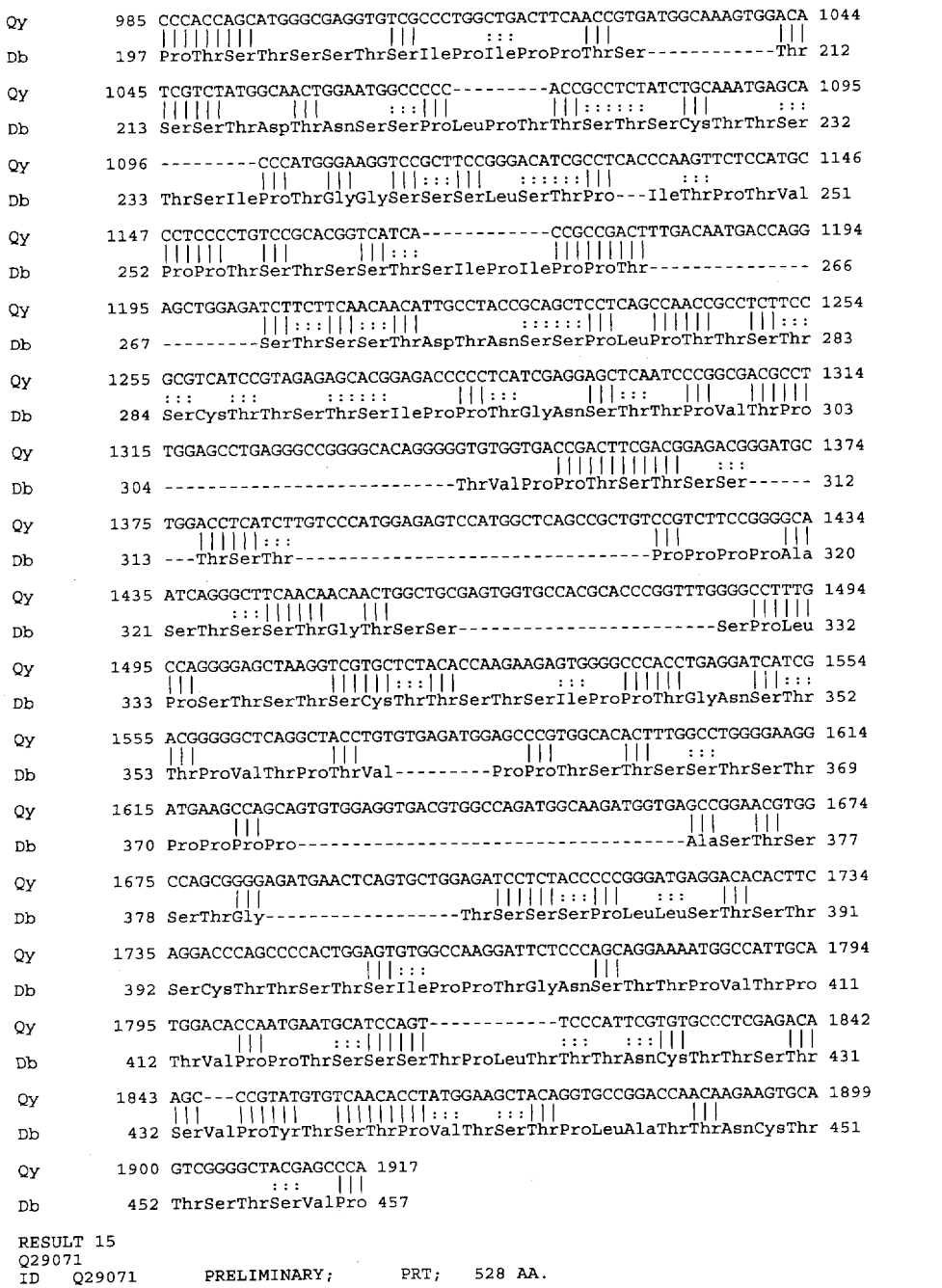
```
Db        401 ThrThrProArgSerThrThrThrThrSerThrSerGlyProThr---ThrThrThrPro 419

Qy        967 GTGCTGGTGTGGACGACC------------CCCACCAGCATGGGCGAGGTGTCGCCCTGG 1014

Db        420 ArgSerThrThrThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThr 439

Qy       1015 CTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACC 1074

Db        440 ThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThrThrSerThr 459

Qy       1075 GCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCG---CCTCAC 1131

Db        460 SerGlyProThrThrThrThrProArgSerThrThrThrThrSerThrSerGlyProThr 479

Qy       1132 CCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACC 1191

Db        480 ThrThrThrProArgSerThrThrThrThrSerThrSerGlyProThrThrThrThrPro 499

Qy       1192 AGGAGCTGGAGATCTTCTTCAACAACATTGCCT---ACCGCAGCTCCTCAGCCAACCGCC 1248

Db        500 ArgSerThrThrThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThr 519

Qy       1249 TCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCA--------- 1299

Db        520 ThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThrThrThrSerThr 539

Qy       1300 ---------ATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGA 1350

Db        540 SerGlyProThrThrThrThrProArgSer---------------------------- 549

Qy       1351 CCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTC 1410

Db        550 ThrThrThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThrThrThr 569

Qy       1411 AGCCGCTGT---CCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGG 1467

Db        570 SerThrSerGlyProThrThrThrThrProArgSerThrThrThrThrSerThrSer--- 588

Qy       1468 TGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGA 1527

Db        589 -----------------GlyProThrThrThrThrProArgSerThrThrThrSer 602

Qy       1528 AGAGTGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGC 1587

Db        603 ThrSerGlyProThr------ThrThrThrProArgSerThrThrThrThrSerThrSer 620

Qy       1588 -----------------CCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTG 1629

Db        621 GlyProThrThrThrThrProArgSerThrThrThrThrSerThrSerGlyProThrThr 640

Qy       1630 TGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGA 1689

Db        641 ------------------------------------ThrThrProArgSerThrThr 647

Qy       1690 ACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCAC 1749

Db        648 Thr-----------ThrSerThrSerGlyProThrThr---------ThrThrProArg 660

Qy       1750 TGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAAT 1809

Db        661 ---------SerThrThrThrThrSerThrSerGlyProThrThrThrThrProArgSer 677

Qy       1810 GCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAA 1869

Db        678 ThrThrThrThrSerThr-----------SerGlyProThrThrThrThrProArgSer 693

Qy       1870 GCTACA---------GGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACG 1920

Db        694 ThrThrThrThrSerThrSerGlyProThrThrThrThrProArgSerThrThrThrThr 713

Qy       1921 AGGATGGCACAGCCT 1935
```

```
Db        714 SerThrSerGlyPro 718

RESULT 12
Q7ULU0
ID   Q7ULU0      PRELIMINARY;      PRT;  1081 AA.
AC   Q7ULU0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to ASPIC.
GN   RB9289.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294149; CAD76179.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1081 AA;  120188 MW;  45592F181A202999 CRC64;

Alignment Scores:
Pred. No.:                4.79e-08      Length:          1081
Score:                    264.00        Matches:         133
Percent Similarity:       38.85%        Conservative:    69
Best Local Similarity:    25.58%        Mismatches:      218
Query Match:              6.51%         Indels:          100
DB:                       16            Gaps:            24

US-09-914-958B-35 (1-2177) x Q7ULU0 (1-1081)

Qy        282 CTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGATGGG------GACTTT------ 329

Db        571 LeuGlyGlyGlyValAlaCysLeuAspPheAspAlaAspGlyArgValAspPheTyrPhe 590

Qy        330 ---------------GAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAG 374

Db        591 AsnGlnAlaAlaGlyGluAlaThrThrArgSerGlyValLysProAsnAlaLeu------ 608

Qy        375 TATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTAC 434

Db        609 TyrArgGlnLeuAspGlyLysPheThrAsnIle---ValAsp------------------ 621

Qy        435 TACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGG 494

Db        622 ValSerMetSerAspAspArgGlyTyrGlyHisGlyValThrAlaGlyAspTrpAsnGln 641

Qy        495 GACGGCCGGGAGGAGATCTACTTCCTCAAC---ACCAATAATGCCTTCTCGGGGGTGGCC 551

Db        642 AspGlyTrpProAspLeuLeuLeuAlaAsnPheGlyGluAsnAlaLeu------------ 657

Qy        552 ACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGAT 611

Db        658 ---------------LeuIleAsnGlnAlaAspGlyThrPheGluLysHisLeuLeuSer 672

Qy        612 GAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTG 671

Db        673 GluLeuSerGluThrGluSerValSerSerGluSerPheThrLeuSerAlaAlaIleAla 692

Qy        672 GACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTG 731

Db        693 AspValThrGlyAspGlnLeuProAspIleValGluValAsnTyr------------Val 708

Qy        732 GGCCCTGATGCCCTC---------ATTGAAATGGACCCTGAGGCCAGTGACCTCTCC--- 779
```

```
Db        709 AspAspSerAlaValArgLysProIleGluArgThrProAspGlyThrProValLysLeu 728

Qy        780 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 839
              |||   |||     :::         :::|||      ||||||          |||
Db        729 ProGlyProLeuHisPheGlnProAlaLeuSerGluValLeuValSerAsnGlyLysGly 748

Qy        840 ------------------------------------------------GGCCGAGGCGTC 851
                                                             |||   ||||||
Db        749 AlaGlyArgArgGlnSerLeuGlyAspSerAlaGlnSerArgSerThrGlyMetGlyVal 768

Qy        852 AGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGG 911
              :::     :::                 ::::::||||||      |||:::
Db        769 MetIleAlaAspMetAspGlnGlnProGlyAsnGluIlePheValAlaAsnAspGlnArg 788

Qy        912 CCTAACTTCCTTTTCCACAACCGGGGCGATGGCACC------------TTTGTGGACGCT 959
              |||     :::           ||||||                  :::   :::
Db        789 AlaAsnHisPheTrpPheGlnAlaGlyAspAlaAsnGluGluGlnGluTrpArgAsnGln 808

Qy        960 GCGGCCAGTGCTGGTGTG------------GACGACCCCCACCAGCATGGGCGAGGTGTC 1007
              ||||||      ||||||                                  |||:::
Db        809 AlaAlaIleArgGlyValAlaValGlyProGlyGlyAspLysAsnAlaCysMetGlyIle 828

Qy       1008 GCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGC 1067
              |||:::||||||||||||||||||    |||:::|||:::          |||:::
Db        829 AlaMetAlaAspPheAsnArgAspGlnLysLeuAspLeuHisValSerAsnPheTyrAsp 848

Qy       1068 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 1127
                    :::       ::::::   |||           |||   :::|||
Db        849 GluTrpAlaAsnHisPheCysGlnThrSerAlaGlyMetPheAlaAspArgAlaValAla 868

Qy       1128 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACG------------GTCATCACCGCC 1175
                       :::   |||      ||||||          |||
Db        869 PheGlyIleAspGlnLeuThrSer------ArgThrThrGlyPheGlyValGlnGlyLeu 886

Qy       1176 GACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCC 1235
              |||:::||||||:::      :::::::::          |||
Db        887 AspTyrAspAsnAsnGlyTrpIleAspLeuLeuValAlaAsnGlyHisValGluAspLeu 906

Qy       1236 TCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAG 1295
              :::                          :::   |||   |||         :::
Db        907 Thr-----------------------------AspArgGlyThrProPheArgMetLys 916

Qy       1296 ------------CTCAATCCCGGCGACGCCTTGGAG-----------------CCTGAG 1325
                        :::|||   |||       |||                   |||:::
Db        917 ThrGlnLeuLeuValAsnGlnGlyLysArgPheGluLeuAlaLysMetAspAspProAsp 936

Qy       1326 GGCCGG-----------------GGCACAGGGGGTGTGGTGACCGACTTCGACGGAGAC 1367
              |||:::                |||  |||           |||||||||   |||
Db        937 GlyGlnTrpGluArgGluArgLeuGlyArgGlyValAlaThrCysAspPheAspArgAsp 956

Qy       1368 GGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTC 1427
              |||   ||||||   ::::::::::::   :::|||   |||:::         :::
Db        957 GlyArgLeuAspAlaValValThrTyrAlaAspSerAlaAlaGlu---------LeuLeu 973

Qy       1428 CGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTG---GTGCCACGCACCCGGTTT 1484
              |||                 |||::::::::::::   :::     ||||||
Db        974 ArgAsnGluThrGluAlaValGlyAsnTyrIleGlnLeuArgLeuIleGlyThrArgSer 993

Qy       1485 GGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTG 1544
                    |||   ||||||   :::   |||   |||
Db        994 GluArgAspAlaIleGlyAlaHisIleGluLeu---ThrAlaGlyAspGlnValTrpThr 1012

Qy       1545 AGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGC 1604
                 ::::::      |||   ||||||   |||       |||   :::   |||||||||
Db       1013 SerValValThrAlaGlyAspGlyTyrAlaCysArgAsnGluAlaIleLeuHisPheGly 1032

Qy       1605 CTGGGGAAGGATGAAGCCAGC------AGTGTGGAGGTGACGTGGCCAGATGGCAAGATG 1658
              ||||||   |||   |||:::      |||:::   |||   ||||||   |||::::::
Db       1033 LeuGly---AspValAlaThrLeuAspSerIleAlaValGlnTrpProSerGlyGluVal 1051

RESULT 13
Q7UPI5
ID   Q7UPI5      PRELIMINARY;      PRT;  1028 AA.
AC   Q7UPI5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Probable acid secreted protein ASPIC.
GN   RB6914.
OS   Rhodopirellula baltica.
OC   Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC   Planctomycetaceae; Pirellula.
OX   NCBI_TaxID=117;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1;
RX   MEDLINE=22735913; PubMed=12835416;
RA   Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA   Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA   Schlesner H., Amann R., Reinhardt R.;
RT   "Complete genome sequence of the marine planctomycete Pirellula sp.
RT   strain 1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR   EMBL; BX294145; CAD75077.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1028 AA;  114349 MW;  C06C31757EFFDF8B CRC64;

Alignment Scores:
Pred. No.:                  7.77e-08        Length:          1028
Score:                      260.50          Matches:         139
Percent Similarity:         40.15%          Conservative:    79
Best Local Similarity:      25.60%          Mismatches:      223
Query Match:                6.42%           Indels:          103
DB:                         16              Gaps:            28

US-09-914-958B-35 (1-2177) x Q7UPI5 (1-1028)

Qy        261 TATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCATGAT 320
              |||:::|||           |||      ||||||   |||  |||  |||  |||
Db        532 TyrGluSer-----------LeuGlyGlyGlyValGlyValIleAspTyrAspAlaAsp 547

Qy        321 GGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTATGAC 380
              |||:::   ::::::      |||    :::                  |||:::
Db        548 GlyAsnProAspLeuTyrPheAlaGlnGlySerAlaAspProProGluLeuArgGlyThr 567

Qy        381 CGGGCCCAGAAGCGGCTGGTGAAC--------------ATCGCGGTCGATGAGCGCAGC 425
              |||      :::          |||             :::||||||||:::|||
Db        568 ArgProAsnGlnLeuPheArgAsnLeuGlyLysArgPheValAlaValGluGlnGlnSer 587

Qy        426 TCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGAC 485
                                  |||      |||   :::   |||::::::      |||
Db        588 Gln---------------CysAspGluPheGlyTyrSerGlnGlyIleSerValGlyAsp 602

Qy        486 ATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAAC------ACCAATAATGCCTTC 539
              ::::::   ||||||     ::::::          |||      |||      |||
Db        603 ValAsnGlnAspGlyTrpProAspLeuLeuValAlaAsnIleGlyLeuAsnArgLeuPhe 622

Qy        540 TCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGAC 599
                       |||      |||                |||||
Db        623 IleAsnCysAlaAspGlyThr---------------PheArg------------------ 631

Qy        600 ATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCT 659
                    |||      :::       |||          |||          |||
Db        632 ---------AspAlaThrAspGlnMetHisGlyGlnAspGluArgPheThrSer---Ser 647

Qy        660 GTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTAC--- 716
              :::|||      |||      |||                  ::::::      ||||||
Db        648 MetAlaIleAlaAspIleAspGlyAspAlaLeuProAspLeuPheGluValAsnTyrIle 667

Qy        717 --------------------------GCCTACGGTAATGTGGGCCCTGATGCCCTC--- 746
                                        :::   :::     |||   :::   |||
Db        668 GlnLeuGluGlyAlaPheAspProProGlnPheAspSerGlnGlyArgGluLeuLeuGln 687

Qy        747 ------ATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGA 800
                    :::       ||||||:::           ||||||
Db        688 GlyProLeuSerGlnLeuProGluSerAspHisTrpTyrArgGlyAspGlyArgGlyAsn 707
```

```
Qy        801 GATGTGGCTGCTGAG------GCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGC 854
Db        708 PheValGlySerGluIleAspGluSerIleAlaLysAlaGlyThrGlyLeuGlyIleVal 727

Qy        855 GTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCT 914
Db        728 IleThrAsnIleAspGlyArgValGlyAsnAspValPheValGlyAsnAspAlaArgPro 747

Qy        915 AACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGT 974
Db        748 AsnHisLeuLeuThrThrLys---GluGlyHisTrpLysAspIleAlaSerIleGlnGly 766

Qy        975 GTGGACGACCCCCACCAGCATGGGCGA----------GGTGTCGCC 1010
Db        767 Val------------AlaSerGlyArgPheGlyLeuSerThrAlaCysMetGlyIleAla 782

Qy       1011 CTGGCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGG---AATGGC 1067
Db        783 ThrGlyAspPheAsnArgAspGlyArgPheAspMetHisValSerAsnPheIleAsnGlu 802

Qy       1068 CCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCC 1127
Db        803 TyrAspAsnLeuPheLeuGln------AsnGluGlyGlyThr---PheArgAspValAla 819

Qy       1128 TCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATC----------ACC 1172
Db        820 ---ProSerTyrGlnLeuProArgPheCysSerAsnAsnValGlyPheGlySerLysSer 838

Qy       1173 GCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTT---------- 1210
Db        839 LeuAspLeuAspArgAspGlyTrpLeuAspValMetThrThrAsnGlyHisIlePheAsp 858

Qy       1211 CAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGA 1270
Db        859 GlnSerHis---LeuAspGluProPheArgGlnLysProLeuLeuIleHisCysCysLys 877

Qy       1271 GCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCG 1330
Db        878 AspArgPheGluArgAlaArgValSerAlaSer-SerAspTyrPheSerGlnGluHisLe 897

Qy       1331 GGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTC 1390
Db        897 uGlyArgGlyLeuAlaLysLeuAspTrpAsnGlnAspGlyArgValAspPheValValTh 917

Qy       1391 CCATGGAGAGTCCATGGCTCAGCCGCTG----------TCCGTCTTCCGGGGCAA 1435
Db        917 rHisLeuAspArgProThrAlaLeuGluLeuAspThrAspThrIleGlyArgGlyLe 937

Qy       1436 TCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGC 1495
Db        937 uGln----------LeuGluLeuIle---GlyValGlnSerGluArgAspAl 950

Qy       1496 CAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGA 1555
Db        950 aIleGlyThrArgValThrIle----------LysAlaAlaHisGlySerLeuValAs 966

Qy       1556 C---------GGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCT 1606
Db        966 pTrpValThrAlaGlyAspGlyTyrLeuCysThrAspGluAlaLeuAspPheGlyLe 986

Qy       1607 GGGGAAG---GATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGC---AAGATGGT 1660
Db        986 uAlaAspCysAspArgValAspGluIleArgValGlnTrpProSerGlyAlaThrGlnTh 1006

Qy       1661 GAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAG---------ATCCTCTA 1711
Db       1006 rPheArgAspIleAlaThrGlyArgArgTyrLeuIleValGluAsnGluSerValLeuHi 1026

Qy       1712 CCCCCGG 1718
Db       1026 sGlnArg 1028
```

```
RESULT 14
Q8TFG4
ID   Q8TFG4      PRELIMINARY;      PRT;   800 AA.
AC   Q8TFG4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative glucoamylase, similar to S. cerevisiae STA1, similar to S.
DE   pombe SPBPJ4664.02.
GN   SPAPB18E9.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   O'Neil S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL691402; CAD27470.1; -.
DR   GeneDB_SPombe; SPAPB18E9.04c; -.
SQ   SEQUENCE   800 AA;  79351 MW;  F65488E43FE7089A CRC64;

Alignment Scores:
Pred. No.:                1.46e-07      Length:          800
Score:                    255.50        Matches:         132
Percent Similarity:       36.31%        Conservative:    59
Best Local Similarity:    25.10%        Mismatches:      180
Query Match:              6.30%         Indels:          155
DB:                       3             Gaps:            24

US-09-914-958B-35 (1-2177) x Q8TFG4 (1-800)

Qy        412 TCGATGAGCGCAGCTCACCCT----------ACTACGCGCTGC------GGGACCGGC 453
Db         63 SerLeuSerThrGlySerProValGluIleThrSerThrSerCysThrThrAspThrSer 82

Qy        454 AGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCT 513
Db         83 AlaSerThrProIleIleThrGluSerThrSerSerThrSerSerAlaSerThrThrGly 102

Qy        514 ACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCA 573
Db        103 SerSerSerSerProLeu----------ProSerThrSerThrSerCysThr 116

Qy        574 AGTTCCGCAATA---------ACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
Db        117 ThrSerThrSerIleProProThrGlyGlySerSerSer---------- 129

Qy        625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
Db        129 ---------- 129

Qy        685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
Db        130 ----------LeuSerThr---ProIleThrProThrVal---------- 138

Qy        745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
Db        139 ----------ProProThrSer---------- 142

Qy        805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
Db        143 ----------ThrSerSerThrSerIleProIleProProThrSerThrSerSerThr 158

Qy        865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
Db        159 AspThrAsnSerAsnProLeuProThrThrSerThrSerCysThrThrSerThrSerIle 178

Qy        925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
Db        179 ProProThrGlyGlySerSerSerLeuSerThr------ProIleThrProThrValPro 196
```



```
RESULT 15
Q29071
ID   Q29071      PRELIMINARY;      PRT;   528 AA.
AC   Q29071;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Gastric mucin (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Yorkshire/Chester White/Hampshire cross; TISSUE=Gastric mucosa;
RX   MEDLINE=95275264; PubMed=7755593;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA   LaMont J.T.;
RT   "Isolation and characterization of cDNA clones encoding pig gastric
RT   mucin.";
RL   Biochem. J. 308:89-96(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Yorkshire/Chester White/Hampshire cross; TISSUE=Gastric mucosa;
RX   MEDLINE=94102478; PubMed=7506218;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA   LaMont J.T.;
RT   "Pig gastric mucin: isolation and characterization of a cDNA clone
RT   with a novel tandem repeat.";
RL   Gastroenterology 106:200-200(1994).
DR   EMBL; U10281; AAC48526.1; -.
DR   PIR; I47141; I47141.
FT   NON_TER       1      1
FT   NON_TER     528    528
SQ   SEQUENCE   528 AA;  49907 MW;  0BF0F6879203B2EA CRC64;

Alignment Scores:
Pred. No.:                3.21e-07      Length:              528
Score:                    249.00        Matches:             156
Percent Similarity:       36.90%        Conservative:        89
Best Local Similarity:    23.49%        Mismatches:          241
Query Match:              6.14%         Indels:              178
DB:                       6             Gaps:                23

US-09-914-958B-35 (1-2177) x Q29071 (1-528)
```

```
Qy        496 ACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGT 555
              :::          |||:::   |||   |||         :::
Db        115 SerValGlnProSerSerSerGlySerValProThrThrSerAlaThrSer--------- 131

Qy        556 ACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGG 615
                 :::|||   ||||||||||||   |||         ||||||
Db        132 ValGlnSerSerSerSerSerSerAlaProThrThrSerAlaThrSer------------ 147

Qy        616 TCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACA 675
                       |||   |||:::|||
Db        148 --------------ValGlnProSerSer---------------------------- 152

Qy        676 GAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCC 735
                            |||:::   ||||||:::   ||||||      :::
Db        153 --------------------SerSerSerProProIleSerSerThrVal------Ser 163

Qy        736 CTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGC 795
              :::   ||||||
Db        164 ValGlnProSer---------------------------------------------- 167

Qy        796 TCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCG 855
                                       |||::::::                |||::::::
Db        168 ------------------------SerSerSerSerAlaProThrThrSerAlaThrSer 179

Qy        856 TGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCT--------------GCGACA 900
                   |||||||||::::::   |||   |||||||||                :::
Db        180 ValGlnProSerSerSerSerSerProProIleSerSerThrValSerValGlnThrSer 199

Qy        901 ATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCTG 960
                     :::   :::   |||||||||                          :::
Db        200 SerSerSerSerValProThrThrSerThrThrSerValGlnProSerSerSerSerSer 219

Qy        961 CGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACT 1020
                 |||            |||:::      |||                |||    :::
Db        220 ValProThrThrSerAlaThrSerValArgSerSerSerSerSerSerThrProIlePro 239

Qy       1021 TCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCT 1080
              ||||||       :::   :::||||||   :::         |||||||||   |||
Db        240 SerThrThrSerValGlnProSerSerSerSerSer---------AlaProThrThrSer 256

Qy       1081 ATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCT 1140
                    |||          |||:::|||   ::::::|||   ||||||:::
Db        257 AlaThrSerValGlnPro---------SerSerSerSerSerThrProIleProSerThr 273

Qy       1141 CCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTGG 1200
                    |||:::   ||||||   ||||||   :::   |||
Db        274 ThrSerValGlnProSerSerSerSerAlaProThrThrSerAlaThrSerValGln 293

Qy       1201 AGATCTTCTTCAACAACATTGCCT---------ACCGCAGCTCCTCAGCCAACCGCCTCT 1251
                 |||||||||::::::   |||         |||   :::   ||||||::::::|||
Db        294 ProSerSerSerSerSerProProIleSerSerThrIleSerValGlnProSerSerSer 313

Qy       1252 TCCGCGTCATCCGTAGAGAGCACGGAGACC------CCCTCATCGAGGAGCTCAATCCCG 1305
              |||:::|||         ||||||   :::      |||||||||       |||   |||
Db        314 SerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySerAlaPro 333

Qy       1306 GCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAG 1365
                 |||                                          ::::::|||
Db        334 ThrThr-------------------------------------------SerAlaThrSer 339

Qy       1366 ACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCT 1425
                       :::||||||   |||       |||       :::|||       ||||||
Db        340 ValGlnProSerSerSerSerSerProProIleSerSerThrIleSerValGlnProSer 359

Qy       1426 TCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTG 1485
              |||   :::          :::|||:::|||                    :::   |||
Db        360 SerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGlySer 379

Qy       1486 GGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGA 1545
                 |||                        ::::::      |||   |||:::
Db        380 AlaPro--------------------ThrThrSerAlaThrSerValGlnProSerSer 392

Qy       1546 GGATCATCGACG---GGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTG 1602
                 ||||||                ||||||   ||||||   |||
Db        393 SerSerSerValProThrThrSerAlaThrSerValArgSerSerSer------------ 408

Qy       1603 GCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGA 1662

Db        408 ------------------------------------------------------------ 408

Qy       1663 GCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATG 1722
              :::   :::   |||           |||           ||||||:::      :::
Db        409 SerSerSerThrProIleProThrThrThrSerValGlnProSerSerSerSerSerVal 428

Qy       1723 AGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAA 1782
                 |||         |||           ||||||         |||   |||
Db        429 ProThrThrSerAlaThr-----------SerValGlnThrSerSerSerSerSerThr 444

Qy       1783 ATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGT------------ 1830
                       :::   |||   :::       |||||||||   |||
Db        445 ProIleProSerThrThrSerValGlnProSerSerSerSerAlaProThrThrSer 464

Qy       1831 ------GCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACA---GGTGCC 1881
                    :::::::::   |||           |||   |||:::   :::|||
Db        465 AlaThrSerValGlnProSerSerSerSerSerProProIleSerSerThrIleSerVal 484

Qy       1882 GGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGG 1941
                 |||:::   |||:::         |||:::   |||         ||||||:::
Db        485 GlnProSerSerSerSerSerProThrThrSerThrThrSerValGlnProSerSer 504

Qy       1942 GCTGGTGGAGCC 1953
              :::|||   |||
Db        505 SerGlySerAla 508
```

Search completed: August 3, 2004, 19:47:58
Job time : 142 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 9, 2004, 17:08:04 ; Search time 5087 Seconds
(without alignments)
12779.628 Million cell updates/sec

Title: US-09-914-958B-35
Perfect score: 2177
Sequence: 1 cgggaggctcgaggccagcc.....tcctcttgaaaaaaaaaaaa 2177

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1544.2	70.9	2089	11	AK049801	AK049801 Mus muscu
	2	1539.4	70.7	2019	11	AK083221	AK083221 Mus muscu
	3	1483.6	68.1	2483	11	AK032328	AK032328 Mus muscu
	4	1438	66.1	2147	11	BC042687	BC042687 Mus muscu
	5	890.8	40.9	1022	13	BX460454	BX460454 BX460454
	6	798	36.7	841	12	BG708845	BG708845 602673155
	7	790.8	36.3	904	13	BX451456	BX451456 BX451456
	8	773.8	35.5	781	12	BI550776	BI550776 603195670
	9	773.6	35.5	1101	13	BQ067249	BQ067249 AGENCOURT
	10	715.8	32.9	770	12	BI758703	BI758703 603024078
	11	708.4	32.5	804	12	BI458314	BI458314 603199270
	12	692.2	31.8	860	12	BI819378	BI819378 603034549
	13	682.6	31.4	926	10	BF527664	BF527664 602040644
	14	655	30.1	1063	12	BI596960	BI596960 603242976
c	15	642.4	29.5	656	13	BQ183029	BQ183029 UI-H-EU0-
	16	634.6	29.2	735	12	BG911024	BG911024 602813023
	17	628.8	28.9	636	10	BF727258	BF727258 by19a09.y
	18	600.4	27.6	713	9	AI792073	AI792073 ag62d02.y
	19	590.4	27.1	740	12	BM948537	BM948537 UI-M-EG0p
	20	574.8	26.4	578	14	CD676152	CD676152 fs32g06.y
	21	574.4	26.4	731	13	BQ178121	BQ178121 UI-M-ER0-
	22	570.6	26.2	580	12	BI549399	BI549399 603190054
	23	566	26.0	567	12	BM707643	BM707643 UI-E-CI1-
	24	556.8	25.6	608	9	AU123536	AU123536 AU123536
	25	556.6	25.6	729	12	BG818593	BG818593 602779443
	26	548.4	25.2	551	12	BM699706	BM699706 UI-E-DW1-
	27	547.8	25.2	622	12	BI824952	BI824952 603033366
	28	541.4	24.9	749	12	BG976852	BG976852 602845756
	29	537.4	24.7	540	14	CB153273	CB153273 K-EST0210
	30	537.4	24.7	679	10	BB429618	BB429618 BB429618
	31	534.2	24.5	539	14	CB152752	CB152752 K-EST0210
	32	532.4	24.5	566	12	BG721807	BG721807 602695986
	33	526.8	24.2	549	12	BM706915	BM706915 UI-E-CQ0-
	34	525.2	24.1	622	12	BG716170	BG716170 602677669
	35	524.4	24.1	721	14	CK221925	CK221925 700248663
c	36	505.4	23.2	754	14	CK221936	CK221936 702812920
	37	498.2	22.9	559	10	BF080090	BF080090 230847 MA
	38	487.6	22.4	1418	12	BM805494	BM805494 AGENCOURT
	39	486	22.3	598	13	BQ637467	BQ637467 he10e02.y
	40	477	21.9	511	9	AI791928	AI791928 ag11c02.y
	41	474.4	21.8	489	10	BF725127	BF725127 bx12e04.y
	42	470.6	21.6	554	10	BF080091	BF080091 230848 MA
	43	467.4	21.5	508	10	BF088585	BF088585 RC1-HT088
	44	462.2	21.2	661	13	BY741935	BY741935 BY741935
	45	459	21.1	660	10	BB653351	BB653351 BB653351

ALIGNMENTS

RESULT 1
AK049801
LOCUS AK049801 2089 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone:C530050M20 product:ASPIC PRECURSOR homolog [Homo sapiens], full insert sequence.
ACCESSION AK049801
VERSION AK049801.1 GI:26340523
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2089)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .2089
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:C530050M20"
/db_xref="MGI:2417318"
/db_xref="taxon:10090"
/clone="C530050M20"
/tissue_type="spinal cord"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
CDS 122. .2062
/note="unnamed protein product; ASPIC PRECURSOR homolog [Homo sapiens] (SPTR|Q9NQ80, evidence: FASTY, 94.1%ID, 91.6%length, match=1821)
putative"
/codon_start=1
/protein_id="BAC33924.1"
/db_xref="GI:26340524"
/translation="MAPSADPGMVRMALLLLPPLWLLPLTGGSQRAEPMFTAVTNSVL
PPDYDSNPTQLNYGVAVTDVDHDGDFEIVVAGYTGPNLVLKYNRAQNRLVNIAVDERS
SPYYALRDRQGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWED
ILSDDVNVARGVASLFAGRSVACVDRTGSGRYSIYIANYAYGDVGPDALIEMDPEASD
LSRGILALRDVAAEAGVSKYTAGRGVSVGPILSSSASDIFCDNENGPNFLFHNQGNGT
FVDTAASAGVDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSAHGKVRFRD
IASPKFSTPSPVRTVIAADFDNDQELEVFFNNIAYRSSSANRLFRVIRREHGDPLIEE
LNPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFSNNWLRVV
PRTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGRDEASSVEVTW
PDGKMVSRSVASEEMNSVLEILYPQDEDKLQNTAPLECGQGFSQQDNGHCMDTNECIQ
FPFVCPRDKPVCVNTYGSYRCRTNKRCNRGYEPNEDGTACVAQVAFLGGYSSAAFRLS
EPLSQASYLSLGLGLCLQLYAL"
polyA_signal 2071. .2076
/note="putative"
polyA_site 2089
/note="putative"
ORIGIN

Query Match 70.9%; Score 1544.2; DB 11; Length 2089;
Best Local Similarity 88.2%; Pred. No. 3.9e-287;
Matches 1714; Conservative 0; Mismatches 223; Indels 6; Gaps 3;

```
Qy          5 AGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAG 64
              ||||||||||||||||||||||||| |||||||| ||     |||||||||||||| ||||||
Db          1 AGGCTCGAGGCCAGCCCGGGACCAGGGCTGGGCGCTGCAAGGCGGCGGCGCCAGCGGCAG 60

Qy         65 AGGCGGCAGCGAGCGCCCGC-TTCCCACGCCCCTAGGCGGCGGGGC--CGAGAGCGGGAG 121
              |||| ||| |||||||||||  ||| |||||| | |||| |||||     ||||||  |||
Db         61 AGGCAGCATCGAGCGCCCGCGCTCCGACGCCCCCAAGCGGTGGGGCTGAGAGAGCCAGAG 120

Qy        122 GATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTG---CT 178
              |||||||||||||||||||||||| ||||||  |||||||     | |||||||| |   ||
Db        121 GATGGCTCCGAGCGCTGACCCGGGCATGGTCAGGATGGCTTTGCTCCTGCTGCCGCCCCT 180

Qy        179 CTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCAC 238
              ||| |||||||  ||||||||| |||||| |||||||| ||||||||||||||||||||||
Db        181 GTGGCTTCTGCCTCTCACTGGGGGATCCCAGAGGGCTGAGCCCATGTTCACTGCAGTCAC 240

Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
              |||||||| |||| |||||||||||||| || ||||||||||||||||||||| |||||
Db        241 CAACTCAGTCCTGCCCCCTGACTATGACAGCAACCCCACCCAGCTCAACTATGGAGTGGC 300

Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACC 358
              ||| || |||||||||||||||||||||||| |||||||| |||||||||||| || ||
Db        301 AGTGACAGATGTGGACCATGATGGGGACTTCGAGATCGTTGTGGCGGGGTACACCGGCCC 360

Qy        359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
              |||||||||||||||||  | || |||||||| |||||||||| || ||||| || ||
Db        361 CAACCTGGTTCTGAAGTACAATCGAGCCCAGAATCGGCTGGTGAATATTGCGGTGGACGA 420

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
              |||||||||||||||||| || ||| |||||| |||||||||||||| ||||||||||||
Db        421 GCGCAGCTCACCCTACTATGCTCTGAGGGACCGTCAGGGGAACGCCATCGGGGTCACAGC 480

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
              ||| |||||||| |||||||||||||| |||||||||||| |||||||||||| ||||||
Db        481 CTGTGACATCGATGGGGACGGCCGTGAGGAGATCTACTTTCTCAACACCAATAACGCCTT 540

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
              ||| || |||||||| ||||| |||||||||||||| ||||| ||||||||||||||||||
Db        541 CTCAGGTGTGGCCACATACACAGACAAGTTGTTCAAATTCCGAAATAACCGGTGGGAAGA 600

Qy        599 CATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTC 658
              |||||||| ||||| ||||| |||||||||| |||||||||||||||||| ||||||||
Db        601 CATCCTGAGTGATGACGTCAATGTGGCCCGTGGAGTGGCCAGCCTCTTTGCGGGACGCTC 660

Qy        659 TGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGC 718
              |||||||||| |||| | |||||||||| |||||||||||||| ||||| || ||
Db        661 CGTGGCCTGTGTAGACAGGACGGGCTCTGGAAGATACTCTATCTACATAGCCAACTATGC 720

Qy        719 CTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTC 778
              ||| ||| |||||||| |||||||||||||| ||||||||||||||||||||||||| ||
Db        721 CTATGGTGATGTGGGGCCTGATGCCCTCATCGAAATGGACCCTGAGGCCAGTGACCTGTC 780
```

```
Qy        779 CCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGG 838
              |||||| || ||||| ||||| || ||||| |||||||||||||||||||| || ||||
Db        781 CCGGGGGATCCTGGCACTCAGGGACGTGGCGGCTGAGGCTGGGGTCAGCAAGTACACAGC 840

Qy        839 GGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGA 898
              |||||| || |||||||| |||||||| |||||||||||||||||| |||||||||| ||
Db        841 GGGCCGGGGTGTCAGCGTAGGCCCCATTCTCAGCAGCAGTGCCTCAGATATCTTCTGTGA 900

Qy        899 CAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGC 958
              ||| ||||| ||||| |||||||| |||||||||  ||| | || ||||||||||  |
Db        901 CAACGAGAACGGGCCCAACTTCCTCTTCCACAACCAAGGCAACGGTACCTTTGTGGATAC 960

Qy        959 TGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGA 1018
              |||||||||||| || || |||||||| || |||||||| |||||||| |||||||| ||
Db        961 TGCGGCCAGTGCCGGCGTAGACGACCCTCATCAGCATGGCCGAGGTGTGGCCCTGGCAGA 1020

Qy       1019 CTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCT 1078
              |||||||||||  |||||||| |||||||||||||||||||||||| ||||| || |||||
Db       1021 TTTCAACCGTGACGGCAAAGTAGACATCGTCTATGGCAACTGGAACGGCCCACATCGCCT 1080

Qy       1079 CTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTT 1138
              |||||| || |||||  | || |||||||||| || |||||||| || || ||||||||
Db       1081 CTATCTACAGATGAGTGCTCACGGGAAGGTCCGGTTTCGGGACATTGCTTCGCCCAAGTT 1140

Qy       1139 CTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCT 1198
              ||||| |||||||||||||| || |||||  | ||||||||||||||||||||||| ||
Db       1141 CTCCACGCCCTCCCCTGTCCGGACTGTCATTGCTGCCGACTTTGACAATGACCAGGAACT 1200

Qy       1199 GGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGT 1258
              |||  |||||||||||||||||||||| | |||||||||||||| |||||||||| ||
Db       1201 GGAAGTCTTCTTCAACAACATTGCCTACAGAAGCTCCTCAGCCAATCGCCTCTTCCGTGT 1260

Qy       1259 CATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGA 1318
              |||||| || |||||||| ||||| |||||||||||||||||| || |||||| | ||
Db       1261 CATCCGCAGGGAGCACGGGGACCCTCTCATCGAGGAGCTCAATCCTGGTGACGCCCTAGA 1320

Qy       1319 GCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGA 1378
              |||||||||||||| |||||||| || |||||||||||||| || || ||||||||||
Db       1321 GCCTGAGGGCCGGGGTACAGGGGGCGTAGTGACCGACTTCGATGGTGATGGGATGCTGGA 1380

Qy       1379 CCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCA 1438
              |||||||||| |||||||||||||||||||||||| ||||| |||||||||| |||||
Db       1381 TCTCATCTTGTCACATGGAGAGTCCATGGCTCAGCCACTGTCTGTCTTCCGGGGAAATCA 1440

Qy       1439 GGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAG 1498
              |||||||| ||||||||| |||| |||| || |||||||||| ||||||||||||||
Db       1441 GGGCTTCAGCAACAACTGGTTGCGTGTGGTACCTCGCACCCGGTTCGGGGCCTTTGCCAG 1500

Qy       1499 GGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATCGACGG 1558
              ||| || ||||| || |||||||||||||||||||||||| |||||  ||||||| || ||
Db       1501 GGGCGCCAAGGTTGTACTCTACACCAAGAAGAGTGGGGCGCACCTACGGATCATTGATGG 1560

Qy       1559 GGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGATGA 1618
              ||| || |||||||||||||| |||||||| |||||||| ||||||||||    ||||||
Db       1561 GGGTTCCGGCTACCTGTGTGAAATGGAGCCTGTGGCACATTTTGGCCTGGGACGGGATGA 1620

Qy       1619 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGGCCAG 1678
              |||||||||||||||||||||||||||||||||||||||||||||||||| |  ||||||||
Db       1621 AGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGAAGTGTGGCCAG 1680

Qy       1679 CGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTTCAGGA 1738
              || |||||||||||| ||| |||||||||||||||||||| |||||||||| ||||||| |
Db       1681 CGAGGAGATGAACTCGGTGTTGGAGATCCTCTACCCCCAGGATGAGGACAAACTTCAGAA 1740

Qy       1739 CCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGCATGGA 1798
              | |||||||||| |||||| |||||||||||||||||||||||| ||||||||||||||||
Db       1741 CACAGCCCCACTAGAGTGCGGCCAAGGATTCTCCCAGCAGGACAATGGCCATTGCATGGA 1800

Qy       1799 CACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAA 1858
              |||||||||||||||||||||||||| |||||||||||||||| ||||||||||||||||
Db       1801 CACCAATGAATGCATCCAGTTCCCATTTGTGTGCCCTCGAGACAAACCCGTATGTGTCAA 1860

Qy       1859 CACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAA 1918
              |||||||||||||||||||||||||||||||| || |  |||| ||||||||| || |||||
Db       1861 CACCTATGGAAGCTACAGGTGCCGGACCAATAAAAGATGCAATCGGGGCTATGAACCCAA 1920

Qy       1919 CGAGGATGGCACAGCCTGCGTGG 1941
              || || |||||||||||| ||||
Db       1921 TGAAGACGGCACAGCCTGTGTGG 1943
```

RESULT 2
AK083221
LOCUS AK083221 2019 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus adult male hippocampus cDNA, RIKEN full-length enriched library, clone:C630029D04 product:ASPIC PRECURSOR homolog [Homo sapiens], full insert sequence.
ACCESSION AK083221
VERSION AK083221.1 GI:26101131
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2019)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

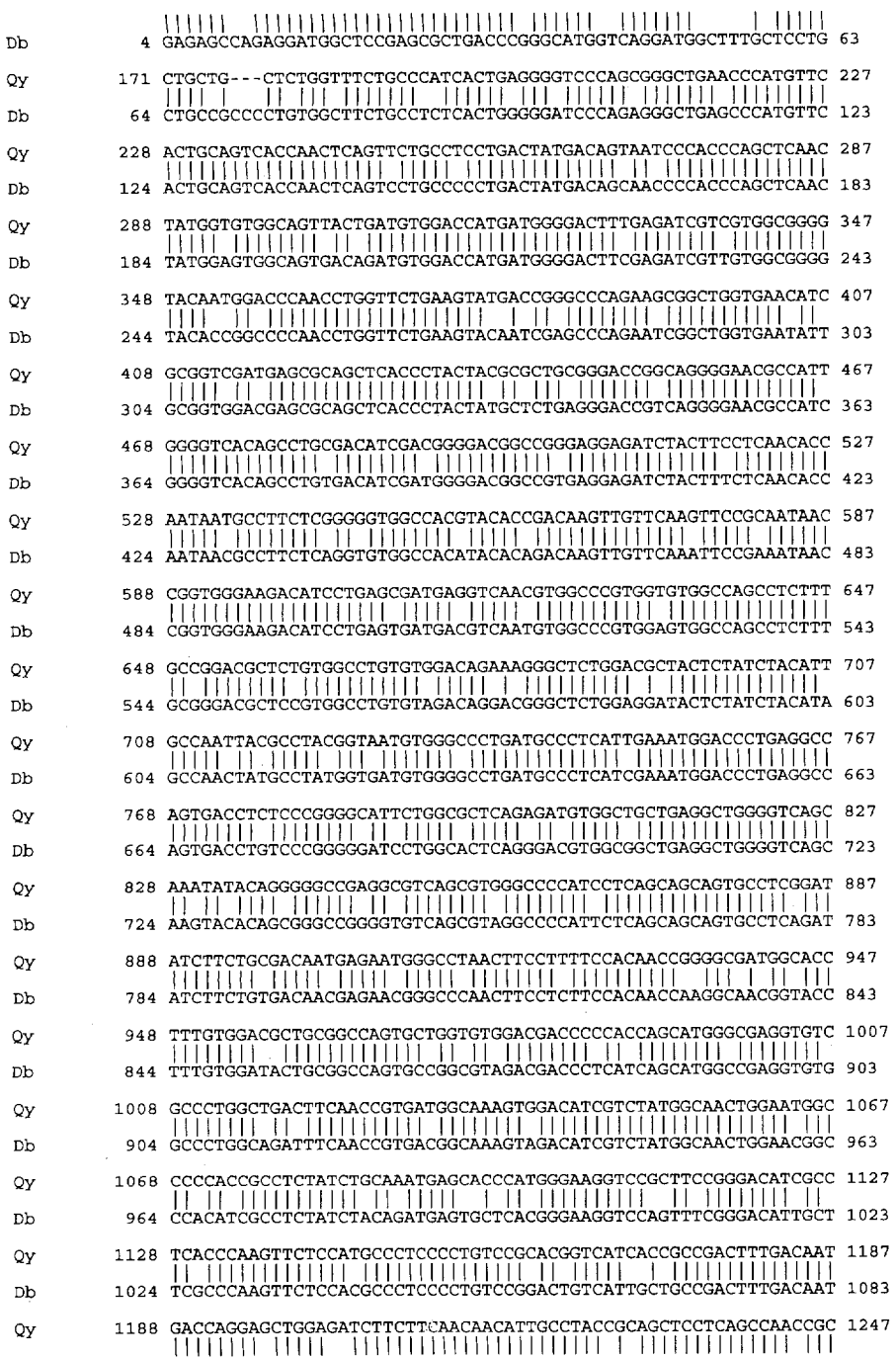
```
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2019
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C630029D04"
                     /db_xref="MGI:2417602"
                     /db_xref="taxon:10090"
                     /clone="C630029D04"
                     /sex="male"
                     /tissue_type="hippocampus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1. .2019
                     /note="ASPIC PRECURSOR homolog [Homo sapiens]
                     (SPTR|Q9NQ80, evidence: FASTY, 94.1%ID, 91.6%length,
                     match=1821)"
ORIGIN

  Query Match             70.7%;  Score 1539.4;  DB 11;  Length 2019;
  Best Local Similarity   88.1%;  Pred. No. 3.2e-286;
  Matches 1722;  Conservative    0;  Mismatches  226;  Indels    7;  Gaps    4;

Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||| ||||||||  ||    ||||||||||||| |||
Db         29 CGGGAGGCTCGAGGCCAGCCCGGGACCAGGGCTGGGCGCTGCAAGGCGGCGGCGCCAGCG 88

Qy         61 GCAGAGGCGGCAGCGAGCGCCCGC-TTCCCACGCCCCTAGGCGGCGGGGC--CGAGAGCG 117
              |||||||| ||| |||||||||||  ||| ||||||| | |||| ||||   ||||||
Db         89 GCAGAGGCAGCATCGAGCGCCCGCGCTCCGACGCCCCCAAGCGGTGGGGCTGAGAGAGCC 148

Qy        118 GGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTG- 176
              |||||||||||||||||||||||| ||||||  |||||||     | |||||||||  |
Db        149 AGAGGATGGCTCCGAGCGCTGACCCGGGCATGGTCAGGATGGCTTTGCTCCTGCTGCCGC 208

Qy        177 --CTCTGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAG 234
                || ||| |||||||  ||||||  ||| ||||||  |||||||  ||||||||||||||||
Db        209 CCCTGTGGCTTCTGCCTCTCACTGGGGGATCCCAGAGGGCTGAGCCCATGTTCACTGCAG 268

Qy        235 TCACCAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCCACCCAGCTCAACTATGGTG 294
              |||||||||||| ||||| |||||||||||||| || |||||||||||||||||||| |
Db        269 TCACCAACTCAGTCCTGCCCCCTGACTATGACAGCAACCCCCACCCAGCTCAACTATGGAG 328

Qy        295 TGGCAGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATG 354
              ||||||| || |||||||||||||||||||||||| |||||||| ||||||||||||| |
Db        329 TGGCAGTGACAGATGTGGACCATGATGGGGACTTCGAGATCGTTGTGGCGGGGTACACCG 388

Qy        355 GACCCAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCG 414
              | |||||||||||||||||||  || |||||||| |||||||||| || ||||| |
Db        389 GCCCCAACCTGGTTCTGAAGTACAATCGAGCCCAGAATCGGCTGGTGAATATTGCGGTGG 448

Qy        415 ATGAGCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCA 474
              | |||||||||||||||||||| || ||| ||||||| |||||||||||||| |||||||
Db        449 ACGAGCGCAGCTCACCCTACTATGCTCTGAGGGACCGTCAGGGGAACGCCATCGGGGTCA 508

Qy        475 CAGCCTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATG 534
              ||||||| |||||||| |||||||||| |||||||||||||| |||||||||||||| |

Db        509 CAGCCTGTGACATCGATGGGGACGGCCGTGAGGAGATCTACTTTCTCAACACCAATAACG 568

Qy        535 CCTTCTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGG 594
              ||||||| || |||||||| ||||| |||||||||||||| ||||| ||||||||||||
Db        569 CCTTCTCAGGTGTGGCCACATACACAGACAAGTTGTTCAAATTCCGAAATAACCGGTGGG 628

Qy        595 AAGACATCCTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGAC 654
              |||||||||||| ||||| ||||| |||||||||| |||||||||||||||| ||||
Db        629 AAGACATCCTGAGTGATGACGTCAATGTGGCCCGTGGAGTGGCCAGCCTCTTTGCGGGAC 688

Qy        655 GCTCTGTGGCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATT 714
              |||| |||||||||| ||||| | |||||||| | |||||||||||||| ||||| |
Db        689 GCTCCGTGGCCTGTGTAGACAGGACGGGCTCTGGAAGATACTCTATCTACATAGCCAACT 748

Qy        715 ACGCCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACC 774
              | ||||| ||| ||||||| |||||||||||||| |||||||||||||||||||||||
Db        749 ATGCCTATGGTGATGTGGGGCCTGATGCCCTCATCGAAATGGACCCTGAGGCCAGTGACC 808

Qy        775 TCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATA 834
              | |||||||| || ||||| ||||| || |||| |||||||||||||||||||| || |
Db        809 TGTCCCGGGGGATCCTGGCACTCAGGGACGTGGCGGCTGAGGCTGGGGTCAGCAAGTACA 868

Qy        835 CAGGGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCT 894
              ||| |||||| || |||||||| |||||||| |||||||||||||||| ||||||||||
Db        869 CAGCGGGCCGGGGTGTCAGCGTAGGCCCCATTCTCAGCAGCAGTGCCTCAGATATCTTCT 928

Qy        895 GCGACAATGAGAA-TGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTG 953
              | ||||| |||||  ||||| |||||||| |||||||||  ||| | || |||||||||
Db        929 GTGACAACGAGAACCGGGCCCAACTTCCTCTTCCACAACCAAGGCAACGGTACCTTTGTG 988

Qy        954 GACGCTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTG 1013
              || |||||||||||||| || || |||||||| || |||||||| |||||| ||||||
Db        989 GATACTGCGGCCAGTGCCGGCGTAGACGACCCTCATCAGCATGGCCGAGGTGTGGCCCTG 1048

Qy       1014 GCTGACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCAC 1073
              || || |||||||||| |||||||| ||||||||||||||||||||| ||||| ||
Db       1049 GCAGATTTCAACCGTGACGGCAAAGTAGACATCGTCTATGGCAACTGGAACGGCCCACAT 1108

Qy       1074 CGCCTCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCC 1133
              |||||||||| || |||||  | || |||||||||| || |||||||| || || |||
Db       1109 CGCCTCTATCTACAGATGAGTGCTCACGGGAAGGTCCGGTTTCGGGACATTGCTTCGCCC 1168

Qy       1134 AAGTTCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAG 1193
              |||||||||| | |||||||||||| || |||||  |||||||||||||||||||||||
Db       1169 AAGTTCTCCACGCCCTCCCCTGTCCGGACTGTCATTGCTGCCGACTTTGACAATGACCAG 1228

Qy       1194 GAGCTGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTC 1253
              || |||||  |||||||||||||||||||||| | |||||||||||||  |||||||||
Db       1229 GAACTGGAAGTCTTCTTCAACAACATTGCCTACAGAAGCTCCTCAGCCAATCGCCTCTTC 1288

Qy       1254 CGCGTCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCC 1313
              || |||||||| || |||||||| ||||| |||||||||||||||||||| || ||||||
Db       1289 CGTGTCATCCGCAGGGAGCACGGGGACCCTCTCATCGAGGAGCTCAATCCTGGTGACGCC 1348

Qy       1314 TTGGAGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATG 1373
              | |||||||||||||||||||||||||| |||||||| || |||||||||||| ||||||
Db       1349 CTAGAGCCTGAGGGCCGGGGTACAGGGGGCGTAGTGACCGACTTCGATGGTGATGGGATG 1408

Qy       1374 CTGGACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGC 1433
              |||||  |||||||||| |||||||||||||||||||||||| ||||| ||||||||||
Db       1409 CTGGATCTCATCTTGTCACATGGAGAGTCCATGGCTCAGCCACTGTCTGTCTTCCGGGGA 1468

Qy       1434 AATCAGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTT 1493
              |||||||||||| |||||||||| ||| |||||  || |||||||||| |||||||||
Db       1469 AATCAGGGCTTCAGCAACAACTGGTTGCGTGTGGTGACCTCGCACCCGGTTCGGGGCCTTT 1528

Qy       1494 GCCAGGGGAGCTAAGGTCGTGCTCTACACCAAGAAGAGTGGGGCCCACCTGAGGATCATC 1553
              |||||||| || |||||| || |||||||||||||||||||||||| |||||  |||||||
Db       1529 GCCAGGGGCGCCAAGGTTGTACTCTACACCAAGAAGAGTGGGGCGCACCTACGGATCATT 1588

Qy       1554 GACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAG 1613
              || |||||  || |||||||||||||| |||||||| |||||||| |||||||||    |
Db       1589 GATGGGGGTTCCGGCTACCTGTGTGAAATGGAGCCTGTGGCACATTTTGGCCTGGGACGG 1648
```

```
Qy       1614 GATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTG 1673
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |  |||
Db       1649 GATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCAAGATGGTGAGCCGAAGTGTG 1708

Qy       1674 GCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACCCCCGGGATGAGGACACACTT 1733
             ||||||| |||||||||||| ||| ||||||||||||||||||| |||||||||| ||||
Db       1709 GCCAGCGAGGAGATGAACTCGGTGTTGGAGATCCTCTACCCCCAGGATGAGGACAAACTT 1768

Qy       1734 CAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCCAGCAGGAAAATGGCCATTGC 1793
             ||| || |||||||||| ||||| ||||||||||||||||||||| ||||||||||||
Db       1769 CAGAACACAGCCCCACTAGAGTGCGGCCAAGGATTCTCCCAGCAGGACAATGGCCATTGC 1828

Qy       1794 ATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCCCTCGAGACAAGCCCGTATGT 1853
             |||||||||||||||||||||||||||||||| ||||||||||||||||| |||||||||
Db       1829 ATGGACACCAATGAATGCATCCAGTTCCCATTTGTGTGCCCTCGAGACAAACCCGTATGT 1888

Qy       1854 GTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGAAGTGCAGTCGGGGCTACGAG 1913
             |||||||||||||||||||||||||||||||| || |  |||| ||||||||| ||
Db       1889 GTCAACACCTATGGAAGCTACAGGTGCCGGACCAATAAAAGATGCAATCGGGGCTATGAA 1948

Qy       1914 CCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTG 1948
             ||||| || || |||||||||| ||||||  | |
Db       1949 CCCAATGAAGACGGCACAGCCTGTGTGGGCCTGGG 1983
```

RESULT 3
AK032328

```
LOCUS       AK032328               2483 bp    mRNA    linear   HTC 18-SEP-2003
DEFINITION  Mus musculus adult male olfactory brain cDNA, RIKEN full-length
            enriched library, clone:6430520G02 product:ASPIC PRECURSOR homolog
            [Homo sapiens], full insert sequence.
ACCESSION   AK032328
VERSION     AK032328.1  GI:26328152
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2483)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .2483
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6430520G02"
                     /db_xref="MGI:2395800"
                     /db_xref="taxon:10090"
                     /clone="6430520G02"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             16. .1956
                     /note="unnamed protein product; ASPIC PRECURSOR homolog
                     [Homo sapiens] (SPTR|Q9NQ80, evidence: FASTY, 94.1%ID,
                     91.6%length, match=1821)
                     putative"
                     /codon_start=1
                     /protein_id="BAC27817.1"
                     /db_xref="GI:26328153"
                     /translation="MAPSADPGMVRMALLLLPPLWLLPLTGGSQRAEPMFTAVTNSVL
                     PPDYDSNPTQLNYGVAVTDVDHDGDFEIVVAGYTGPNLVLKYNRAQNRLVNIAVDERS
                     SPYYALRDRQGNAIGVTACDIDGDGREEIYFLNTNNAFSGVATYTDKLFKFRNNRWED
                     ILSDDVNVARGVASLFAGRSVACVDRTGSGGYSIYIANYAYGDVGPDALIEMDPEASD
                     LSRGILALRDVAAEAGVSKYTAGRGVSVGPILSSSASDIFCDNENGPNFLFHNQGNGT
                     FVDTAASAGVDDPHQHGRGVALADFNRDGKVDIVYGNWNGPHRLYLQMSAHGKVQFRD
                     IASPKFSTPSPVRTVIAADFDNDQELEVFFNNIAYRSSSANRLFRVIRREHGDPLIEE
                     LNPGDALEPEGRGTGGVVTDFDGDGMLDLILSHGESMAQPLSVFRGNQGFSNNWLRVV
                     PRTRFGAFARGAKVVLYTKKSGAHLRIIDGGSGYLCEMEPVAHFGLGRDEASSVEVTW
                     PDGKMVSRSVASEEMNSVLEILYPQDEDKLQNTAPLECGQGFSQQDNGHCMDTNECIQ
                     FPFVCPRDKPVCVNTYGSYRCRTNKRCNRGYEPNEDGTACVAQVAFLGGYSSAAFRLS
                     EPLSQASYLSLGLGLCLQLYAL"
ORIGIN
```

```
  Query Match           68.1%;  Score 1483.6;  DB 11;  Length 2483;
  Best Local Similarity  88.4%;  Pred. No. 2e-275;
  Matches 1622;  Conservative    0;  Mismatches  209;  Indels    3;  Gaps    1;

Qy       111 GAGAGCGGGAGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTG 170
```



RESULT 4
BC042687
LOCUS BC042687 2147 bp mRNA linear HTC 02-JAN-2003
DEFINITION Mus musculus, clone IMAGE:3157049, mRNA.
ACCESSION BC042687
VERSION BC042687.1 GI:27469363
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2147)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 6 Row: p Column: 3
            This clone has the following problem: no 5' EST match.
FEATURES              Location/Qualifiers
     source           1. .2147
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3157049"
                      /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                      ductal carcinoma. 5 month old virgin mouse."
                      /clone_lib="NCI_CGAP_Mam6"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
ORIGIN

  Query Match             66.1%;  Score 1438;  DB 11;  Length 2147;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-266;
  Matches 1438;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 60

Qy        565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 120

Qy        625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 180

Qy        685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 240

Qy        745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 300

Qy        805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 360

Qy        865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 420

Qy        925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 480

Qy        985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 540

Qy       1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 600

Qy       1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 660

Qy       1165 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 1224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCATCACCGCCGACTTTGACAATGACCAGGAGCTGGAGATCTTCTTCAACAACATTGCCT 720

Qy       1225 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 1284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCCC 780

Qy       1285 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 1344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAGCCTGAGGGCCGGGGCACAGGGGGTG 840

Qy       1345 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 1404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TGGTGACCGACTTCGACGGAGACGGGATGCTGGACCTCATCTTGTCCCATGGAGAGTCCA 900

Qy       1405 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 1464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAGGGCTTCAACAACAACTGGCTGCGAG 960

Qy       1465 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TGGTGCCACGCACCCGGTTTGGGGCCTTTGCCAGGGGAGCTAAGGTCGTGCTCTACACCA 1020

Qy       1525 AGAAGAGTGGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 AGAAGAGTGGGGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGG 1080

Qy       1585 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 AGCCCGTGGCACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGC 1140

Qy       1645 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 CAGATGGCAAGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGA 1200

Qy       1705 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 TCCTCTACCCCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAG 1260

Qy       1765 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 GATTCTCCCAGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCAT 1320

Qy       1825 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 TCGTGTGCCCTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGA 1380

Qy       1885 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CCAACAAGAAGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGG 1438

RESULT 5
BX460454
LOCUS       BX460454                1022 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX460454 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF014YH11 5-PRIME, mRNA sequence.
ACCESSION   BX460454
VERSION     BX460454.1  GI:31037049
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
```

```
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            5247.r For more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF014CD06QP1&cluster=5247.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF014CD06QP1.
FEATURES             Location/Qualifiers
     source          1. .1022
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF014YH11"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match             40.9%;  Score 890.8;  DB 13;  Length 1022;
  Best Local Similarity   98.6%;  Pred. No. 3.2e-161;
  Matches  947;  Conservative    3;  Mismatches    5;  Indels    5;  Gaps    5;

Qy         20 CCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGG-CGGCAGAGGCGGCAGCGAGC 78
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         64 CCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGACGGCAGAGGCGGCAGCGAGC 123

Qy         79 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAG-CGGGAGGATGGCTCCGAGCGCT 137
              ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        124 GCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGACGGGAGGATGGCTCCGAGCGCT 183

Qy        138 GACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACT 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 GACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACT 243

Qy        198 GAGGGGTCCCAG-CGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCC 256
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        244 GAGGGGTCCCAGACGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCC 303

Qy        257 TGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCA 316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCA 363

Qy        317 TGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTA 376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 TGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTA 423

Qy        377 TGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTA 436
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        424 TGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA-CTCACCCTACTA 482

Qy        437 CGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGA 496
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        483 CGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGA 542

Qy        497 CGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTA 602

Qy        557 CACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGT 616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 CACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGT 662

Qy        617 CAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAG 676
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 CAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAG 722

Qy        677 AAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCC 736
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 AAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCC 782

Qy        737 TGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCT 796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCT 842

Qy        797 CAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGT 856
              |||||||||||||||||||||||||||||||||||||:||||:||||||||||
Db        843 CAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACARGGGGSCGAGGCGTCACGTR 902

Qy        857 GGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAA 916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 GGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAA 962

Qy        917 CTTCCTTTTCCACAA-CCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTG 975
              ||||||||||||||| ||||||||||||||||||||||||||||||||:||||||||
Db        963 CTTCCTTTTCCACAACCCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCARTGCTGGTG 1022


RESULT 6
BG708845
LOCUS       BG708845                841 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602673155F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4795957 5',
            mRNA sequence.
ACCESSION   BG708845
VERSION     BG708845.1  GI:13986590
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10679  row: m  column: 14
            High quality sequence stop: 822.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4795957"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN
```

```
  Query Match           36.7%;  Score 798;  DB 12;  Length 841;
  Best Local Similarity  97.0%;  Pred. No. 2.3e-143;
  Matches  813;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 603

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 663

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 722
              |||||||||||||||||||||||||||||||||||||||||||| || | | |  |
Db        664 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATGGCAATTACGCCTACG 723

Qy        723 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGG 782
              |   ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        724 GTACATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTACCGG 783

Qy        783 GGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 GGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTACAGAATATACCAGGGG 841


RESULT 7
BX451456
LOCUS       BX451456                904 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX451456 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF014YH11 5-PRIME, mRNA sequence.
ACCESSION   BX451456
VERSION     BX451456.1  GI:31022247
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 904)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5247.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAF012ZE12_AF01150_1&cluster=5247.r.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAF012ZE12_AF01150_1.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF014YH11"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match           36.3%;  Score 790.8;  DB 13;  Length 904;
  Best Local Similarity  96.8%;  Pred. No. 5.9e-142;
  Matches  860;  Conservative    0;  Mismatches   22;  Indels    6;  Gaps    5;

Qy        838 GGGGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCG 897
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 GGGGCCGAGGCGTCAGCGTGGGCCCCATCCT-AGCAGCAGTGCCTCGGATATCTTCTGCG 59

Qy        898 ACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACG 957
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         60 ACAATGAGAATGGGCCTAACTTCCTTTTCC-CAACCGGGGCGATGGCACCTTTGTGGACG 118

Qy        958 CTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTG 1017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 CTGCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTG 178

Qy       1018 ACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCC 1077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ACTTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCC 238

Qy       1078 TCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGT 1137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TCTATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGT 298

Qy       1138 TCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGC 1197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 TCTCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGC 358

Qy       1198 TGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCG 1257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TGGAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCG 418

Qy       1258 TCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGG 1317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 TCATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGG 478

Qy       1318 AGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGG 1377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AGCCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGG 538

Qy       1378 ACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATC 1437
```

```
Db        539 ACCTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATC 598

Qy       1438 AGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTTGGGGCCTTTGCCA 1497
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 AGGGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGGTTTGGGGCCTTGCCA 658

Qy       1498 GGGGAGCTAAGGTCGTGCTCTACACCAAG-AAGAGTGGGGCCCACCTGAGGATCATCGAC 1556
             |||||||||| |||||||||||||||||| ||||||||| |||||||||||||||||||
Db        659 GGGGAGCTAAAGTCGTGCTCTACACCAAGAAAGAGTGGGCCCCACCTGAGGATCATCGAC 718

Qy       1557 GGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTGGCCTGGGGAAGGAT 1616
             |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db        719 GGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGGCACACTTTTGCCTGGGGAAAGAT 778

Qy       1617 GAAGCCAGCAGTGTGGAGGT--GACGTGGCCAGATGGCAAGATGGTGAGCCGGAACGTGG 1674
             |||||||||||||||| ||     |||| |||||| |  |||||||||| | ||||||
Db        779 GAAGCCAGCAGTGTGGAAGTTGACCGTGCCCAGATTGGCAGATGGTGAGCCCGGACGTGG 838

Qy       1675 CCAGCGGGGAGATGAACTCAGTGCTGG-AGATCCTCTACCCCCGGGAT 1721
             |||||||||||| ||||| |||||||| |||  || ||||||||||||
Db        839 CCAGCGGGGAGAAGAACTAAGTGCTGGAAGAATCTTTACCCCCGGGAT 886
```

RESULT 8
BI550776

```
LOCUS       BI550776                781 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603195670F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5275168 5',
            mRNA sequence.
ACCESSION   BI550776
VERSION     BI550776.1  GI:15438088
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11694  row: l  column: 17
            High quality sequence stop: 763.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5275168"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN
```

```
  Query Match           35.5%;  Score 773.8;  DB 12;  Length 781;
  Best Local Similarity   99.7%;  Pred. No. 1.1e-138;
  Matches  775;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAAC 362
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 603

Qy        603 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 662
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CTGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTG 663

Qy        663 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 722
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 GCCTGTGTGGACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTAC 723

Qy        723 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        724 GGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGAACTCTCC 780
```

RESULT 9
BQ067249

```
LOCUS       BQ067249                1101 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6767325 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5751633
            5', mRNA sequence.
ACCESSION   BQ067249
VERSION     BQ067249.1  GI:19896295
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12784  row: i  column: 10
            High quality sequence stop: 612.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5751633"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_115"
                     /note="Organ: pooled brain, lung, testis; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 6 male brains, age range 23-27; 1
                     male lung, age 27; and 1 male testis, age 69.  Library is
                     oligo-dT primed and directionally cloned (EcoRV site is
                     destroyed upon  cloning).  Average insert size 1.8 kb,
                     insert size range 1-3 kb. Library is normalized and
                     enriched for full-length clones and was  constructed by C.
                     Gruber (Invitrogen).  Research Genetics tracking  code
                     021. Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             35.5%;  Score 773.6;  DB 13;  Length 1101;
  Best Local Similarity   93.9%;  Pred. No. 1.3e-138;
  Matches  862;  Conservative    0;  Mismatches   41;  Indels   15;  Gaps    5;

Qy        720 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCTCC 60

Qy        780 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGG 120

Qy        840 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCGAGGCGTCAGCGTGGGCCCCATCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGAC 180

Qy        900 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGCGATGGCACCTTTGTGGACGCT 240

Qy        960 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCGGCCAGTGCTGGTGTGGACGACCCCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGAC 300

Qy       1020 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCGTGATGGCAAAGTGGACATCGTCTATGGCAACTGGAATGGCCCCCACCGCCTC 360

Qy       1080 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATCTGCAAATGAGCACCCATGGGAAGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTC 420

Qy       1140 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 1199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCATGCCCTCCCCTGTCCGCACGGTCATCACCGCCGACTTTGACAATGACCAGGAGCTG 480

Qy       1200 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 1259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAGATCTTCTTCAACAACATTGCCTACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTC 540

Qy       1260 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 1319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        541 ATCCGTAGAGAGCACGGAGACCCCCTCATCGAGGAGCTCAATCCCGGCGACGCCTTGGAG 600

Qy       1320 CCTGAGGGCCGGGGCACAGGGGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGAC 1379
              ||||||||||||     || ||||||||||||||||||||||||||||||||||||||
Db        601 CCTGAGGGCCGGGCACAGGGNGGTGTGGTGACCGACTTCGACGGAGACGGGATGCTGGAC 660

Qy       1380 CTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAG 1439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTCATCTTGTCCCATGGAGAGTCCATGGCTCAGCCGCTGTCCGTCTTCCGGGGCAATCAG 720

Qy       1440 GGCTTCAACAACAACTGGCTGCGAGTGGTGCCACGCACCCGGTTT---GGGGCCTTTGCC 1496
              ||||||||||||   |||||||||||||||||||||||||||| |   ||| | || |
Db        721 GGCTTCAACAACACTGGGCTGCGAGTGGTGCCACGCACCCGGTNTTGGGGGCCTTTGCCA 780

Qy       1497 AGGGGAGCTAAGGTCGTGCTCTACACC-AAGAAGAGTGGGGCCCACC-----TGAGGATC 1550
              |||||||||| |||||||||||||| |||||||||||||||  |      || |||
Db        781 GGGGGAGCTAAAGTCGTGCTCTACACCAAAGAAGAGTGGGGCCCCACCTGAGGGATCATC 840

Qy       1551 ATCGACGGGGGCTCAGGCTACCT---GTGTGAGATGGAGCCCGTGGCACACTTTGGCC-- 1605
                 ||||   |||  |||||   ||| ||   | ||||||||||||||||| | |
Db        841 GAACGGGGGGCTCCAGCTTACCTGTGGTGAGAATGGAAGCCCGTGGCACACTTTTGGCCT 900

Qy       1606 -TGGGGAAGGATGAAGCC 1622
               |||||||||||||||||
Db        901 TGGGGGAAGGATGAAGCC 918


RESULT 10
BI758703
LOCUS       BI758703                 770 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603024078F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194530 5',
            mRNA sequence.
ACCESSION   BI758703
VERSION     BI758703.1  GI:15750294
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 770)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11486  row: l  column: 19
            High quality sequence start: 3
            High quality sequence stop: 770.
FEATURES             Location/Qualifiers
     source          1. .770
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194530"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
ORIGIN
```

```
  Query Match             32.9%;  Score 715.8;  DB 12;  Length 770;
  Best Local Similarity   98.6%;  Pred. No. 1.6e-127;
  Matches  754;  Conservative    0;  Mismatches    7;  Indels    4;  Gaps    3;

Qy         21 CGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGA-GGCGGCAGCGAGCG 79
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db          6 CGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCAGATGGCGGCAGCGAGCG 65

Qy         80 CCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGA--GAGCGGGAGGATGGCTCCGAGCGCT 137
              ||||||||||||||||||||||||||||||||      ||  ||||||||||||||||||
Db         66 CCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGATGAGCGGGATGGATGGCTCCGAGCGCT 125

Qy        138 GACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACT 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGTTTCTGCCCATCACT 185

Qy        198 GAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCT 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACTCAGTTCTGCCTCCT 245

Qy        258 GACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 GACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTACTGATGTGGACCAT 305

Qy        318 GATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACCTGGTTCTGAAGTAT 365

Qy        378 GACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTAC 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCAGCTCACCCTACTAC 425

Qy        438 GCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCGACATCGACGGGGAC 497
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        426 GCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCGACATCGACGGGGAC 485

Qy        498 GGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTAC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 GGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTAC 545

Qy        558 ACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTC 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 ACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTC 605

Qy        618 AACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGA 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 AACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGA 665

Qy        678 AAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCT 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 AAGGGCTCTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCT 725

Qy        738 GATGCCCTCATTGAAATGGACCC-TGAGGCCAGTGACCTCTCCCG 781
              |||||||||||||||||||||| |||||||||||||||||||||
Db        726 GATGCCCTCATTGAAATGGACCCTTGAGGCCAGTGACCTCTCCCG 770


RESULT 11
BI458314
LOCUS       BI458314                804 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603199270F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5278624 5',
            mRNA sequence.
ACCESSION   BI458314
VERSION     BI458314.1  GI:15248970
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11703  row: 1  column: 17
            High quality sequence stop: 785.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5278624"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             32.5%;  Score 708.4;  DB 12;  Length 804;
  Best Local Similarity   97.8%;  Pred. No. 4.4e-126;
  Matches  782;  Conservative    0;  Mismatches   11;  Indels    7;  Gaps    6;

Qy          4 GAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCA 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGCA 61

Qy         64 GAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGGA 121

Qy        124 TGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGGT 181

Qy        184 TTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACT 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 TTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAACT 241

Qy        244 CAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTA 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTTA 301

Qy        304 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACC 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAACC 361

Qy        364 TGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA 423
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        362 TGGTTCTGAAGTATGACCAGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGCA 421

Qy        424 GCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGCG 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GCTCACACTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGCG 481

Qy        484 ACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGG 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 ACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGG 541
```

```
Qy        544 GGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCC 603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 GGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCC 601

Qy        604 TGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCT-CTGTG 662
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        602 TGAGCGATGAGGTCAACGTGGCCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTGCTGTG 661

Qy        663 GCCTGTG-TGGACAGAAAGGGCTC--TGGACGCTACTCTATC-TACATTGCCAATT-ACG 717
              ||||||| ||||| |||||||||||   | ||||||||||| |||||||||||| ||| 
Db        662 GCCTGTGTTGGACCGAAAGGGCTCCTGGACCGCTACTCTATCTTACATTGCCAATTAACG 721

Qy        718 CCTACGGTAATGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGCCAGTGACCTCT 777
              |||||||||||||| |||||||||||||||||   |||||  |||||||||||||||||
Db        722 CCTACGGTAATGTTGGCCCTGATGCCCTCATTTGAATGG-GCCTGAGGCCAGTGACCTCT 780

Qy        778 CCCGGGGGCATTCTGGCGCTC 797
              |||||||||||||||||||||
Db        781 CCCGGGGGCATTCTGGCGCTC 800
```

RESULT 12
BI819378
LOCUS BI819378 860 bp mRNA linear EST 04-OCT-2001
DEFINITION 603034549F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5175469 5', mRNA sequence.
ACCESSION BI819378
VERSION BI819378.1 GI:15930928
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 860)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11437 row: b column: 14
High quality sequence stop: 824.
FEATURES Location/Qualifiers
source 1. .860
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5175469"
/lab_host="DH10B"
/clone_lib="NIH_MGC_115"
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
ORIGIN

Query Match 31.8%; Score 692.2; DB 12; Length 860;
Best Local Similarity 96.1%; Pred. No. 6e-123;
Matches 816; Conservative 0; Mismatches 23; Indels 10; Gaps 10;

```
Qy          1 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 CGGGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCG 72

Qy         61 GCAGAGGCGGCAGCGAGCGCC-CGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGG 119
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         73 GCAGAGGCGGCAGCGAGCGCCGCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGG 132

Qy        120 AGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTC 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 AGGATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTC 192

Qy        180 -TGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCAC 238
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 ATGGTTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCAC 252

Qy        239 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 CAACTCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGC 312

Qy        299 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACC 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 AGTTACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGGTACAATGGACC 372

Qy        359 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 CAACCTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGA 432

Qy        419 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGC 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GCGCAGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGC 492

Qy        479 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 538
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 CTGCGACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTT 552

Qy        539 CTCGGGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGA 598
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        553 CTCGGGGGTGGCCACGTACACCGACAAG-TGTTCAAGTTCCGCAATAACCGGTGGGAAGA 611

Qy        599 CATCCTGAGCGATGA-GGTCAACGTGGCCCGTGGTG-TGGCCAGCCTCTTTGCCGGACGC 656
              |||||||||||||||  |||||||||||||||||  ||| |||||||||||||||||||
Db        612 CATCCTGAGCGATGAGGGTCAACGTGGCCCGGGGTGTTGGCCAGCCTCTTTGCCGGACGC 671

Qy        657 TCT-GTGGCCTGTGTGG-ACAGAAAGGGCTCTGGACGCTACTCTATCTACATTGCCAATT 714
              ||| |||||||||||||| |||||||||| ||||||||||||||||||||||||||||||
Db        672 TCTGGTGGCCTGTGTGGAACAGAAAGGGTTCTGGACGCTACTCTATCTACATTGCCAATT 731

Qy        715 ACG-CCTACGGTAAT-GTGGGCCCTGATGCCCTCATTGAAATGGACCCTGA-GGCCAGTG 771
              ||| |||||||||||| |||||||||||||||||||||||||||||||| ||||||||
Db        732 ACGCCCTACGGTAATGGTGGGCCCTGATGCCCTCATTGAAATGGACCCTGAGGGCCAGTG 791

Qy        772 ACCTCTCCCGGGGCATTCTGGCGCTCAGAGATGTGGCTGCTGAGGCTGGGGTCAGCAAAT 831
              |||||||| |||   |   |||||||||||||||||||||   |   |||||||  |||
Db        792 ACCTCTCCGGGGCATTCTGGGCGCTCAGAGATGTGGCTTGGCTGAGCTGGGTCAGGCAAT 851

Qy        832 ATACAGGGG 840
              |||| || |
Db        852 ATACGGGCG 860
```

RESULT 13
BF527664
LOCUS BF527664 926 bp mRNA linear EST 11-DEC-2000
DEFINITION 602040644F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4178363 5', mRNA sequence.
ACCESSION BF527664
VERSION BF527664.1 GI:11615027
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 926)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9486  row: h  column: 12
            High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1. .926
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4178363"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn67"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             31.4%;  Score 682.6;  DB 10;  Length 926;
  Best Local Similarity   95.5%;  Pred. No. 4.4e-121;
  Matches  746;  Conservative    0;  Mismatches   29;  Indels    6;  Gaps    4;

Qy        505 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCGGGGGTGGCCACGTACACCGACA 60

Qy        565 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATCCTGAGCGATGAGGTCAACGTGG 120

Qy        625 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCCGTGGTGTGGCCAGCCTCTTTGCCGGACGCTCTGTGGCCTGTGTGGACAGAAAGGGCT 180

Qy        685 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGGACGCTACTCTATCTACATTGCCAATTACGCCTACGGTAATGTGGGCCCTGATGCCC 240

Qy        745 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCATTGAAATGGACCCTGAGGCCAGTGACCTCTCCCGGGGCATTCTGGCGCTCAGAGATG 300

Qy        805 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGCTGCTGAGGCTGGGGTCAGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCCA 360

Qy        865 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCAGCAGCAGTGCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTT 420

Qy        925 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCCACAACCGGGGCGATGGCACCTTTGTGGACGCTGCGGCCAGTGCTGGTGTGGACGACC 480

Qy        985 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCCACCAGCATGGGCGAGGTGTCGCCCTGGCTGACTTCAACCGTGATGGCAAAGTGGACA 540

Qy       1045 TCGTCTATGGCAACTGGAATGGCCCCCACCGCCTCTATCTGCAAATGAGCACCCATGGGA 1104
              |||||||||||||||||||||||    |    ||||||||||||||||||||||||||||
Db        541 TCGTCTATGGCAACTGGAATGGCCC---CACGCTCTATCTGCAAATGAGCACCCATGGGA 597

Qy       1105 AGGTCCGCTTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 1164
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 AGGTCCG-TTCCGGGACATCGCCTCACCCAAGTTCTCCATGCCCTCCCCTGTCCGCACGG 656

Qy       1165 TCATCACCGCCGACTTTGACAATGACCAGGAG-CTGGAGATCTTCTTCAACAACATTGCC 1223
              |||       ||||||||||||||||||||||| ||||| |||||||||||||| |||| |
Db        657 TCAATCAACGCGACTTTGACAATGACCAGGAGCCTGGAAATCTTCTTCAACAAAATTG-C 715

Qy       1224 TACCGCAGCTCCTCAGCCAACCGCCTCTTCCGCGTCATCCGTAGAGAGCACGGAGACCCC 1283
              ||||| || || ||||| ||||| |||||||||||||  |  | ||||  ||    |||
Db        716 TACCGAAGTTCTTCAGCAAACCGGCTCTTCCGCGTCAATCCGAAAGAGACCGAGACCCCT 775

Qy       1284 C 1284
              |
Db        776 C 776


RESULT 14
BI596960
LOCUS       BI596960               1063 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603242976F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5285419 5',
            mRNA sequence.
ACCESSION   BI596960
VERSION     BI596960.1  GI:15489899
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1063)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11721  row: g  column: 20
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .1063
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5285419"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             30.1%;  Score 655;  DB 12;  Length 1063;
  Best Local Similarity   88.6%;  Pred. No. 9.6e-116;
  Matches  851;  Conservative    0;  Mismatches   85;  Indels   25;  Gaps   12;

Qy          3 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 62
```

```
Db          4 GGAGGCTCGAGGCCAGCCCGGGACCGGGGCTGGGAGCAAGCAGGCGGCGGCGCCGGCGGC 63

Qy         63 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 122

Db         64 AGAGGCGGCAGCGAGCGCCCGCTTCCCACGCCCCTAGGCGGCGGGGCCGAGAGCGGGAGG 123

Qy        123 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 182

Db        124 ATGGCTCCGAGCGCTGACCCCGGCATGTCCAGGATGTTACCGTTCCTGCTGCTGCTCTGG 183

Qy        183 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 242

Db        184 TTTCTGCCCATCACTGAGGGGTCCCAGCGGGCTGAACCCATGTTCACTGCAGTCACCAAC 243

Qy        243 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 302

Db        244 TCAGTTCTGCCTCCTGACTATGACAGTAATCCCACCCAGCTCAACTATGGTGTGGCAGTT 303

Qy        303 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 362

Db        304 ACTGATGTGGACCATGATGGGGACTTTGAGATCGTCGTGGCGGGGTACAATGGACCCAAC 363

Qy        363 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 422

Db        364 CTGGTTCTGAAGTATGACCGGGCCCAGAAGCGGCTGGTGAACATCGCGGTCGATGAGCGC 423

Qy        423 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATTGGGGTCACAGCCTGC 482

Db        424 AGCTCACCCTACTACGCGCTGCGGGACCGGCAGGGGAACGCCATCGGGGTCACAGCCTGC 483

Qy        483 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 542

Db        484 GACATCGACGGGGACGGCCGGGAGGAGATCTACTTCCTCAACACCAATAATGCCTTCTCG 543

Qy        543 GGGGTGGCCACGTACACCGACAAGTTGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602

Db        544 GGGGTGGCCACGTACACCGACAAG-TGTTCAAGTTCCGCAATAACCGGTGGGAAGACATC 602

Qy        603 CTGAGCGATGAGGTCAA-CGTGGCCCGTGGTGTGGCCAGCCTCTTTG-CCGGACGCTCTG 660

Db        603 CTGAGCGATGAGGTCAACCGTGTGCCGTGGTGTGGCCAGCCTCTTTGNCCGGACGCTCTG 662

Qy        661 TGGCCTGTGTGGACAGAAAGGGCTCTGG-ACGCTACTCTA---TCTACATTGCCAATTAC 716

Db        663 TGGCCTGTGTGGACCGAAAGGGCTCTGGAACGCTACTCTATTCTACATTTGCCAATTTAC 722

Qy        717 GCCTACGGTAATGT----GGGCCCTGATGCCCT-CATTGAAATGGACCCTGAGGCCAGTG 771

Db        723 GCCCTACGTGTAATTGTGGCGCCCTGATGCCCTCCGTTGCAATGGACCCTGAGGGCAGTT 782

Qy        772 AC---CTCTCCCGGGGCATTCTGGCGCTC--AGAGATGTGGCTGCTGA--GGCTGGGGTC 824

Db        783 GCCCTCTACGCGGGGGCATTCTGGCGCTCCAGCGATTGTGGTTGGTGAGCGCCTGGGTCA 842

Qy        825 AGCAAATATACAGGGGGCCGAGGCGTCAGCGTGGGCCCC-----ATCCTCAGCAGCAG-T 878

Db        843 GGCACTATTCACGGGGGCCGAGGCGTACACACGGGGGCCCCGATGTCCTCAGCAGGAGAT 902

Qy        879 GCCTCGGATATCTTCTGCGACAATGAGAATGGGCCTAACTTCCTTTTCCACAACCGGGGC 938

Db        903 GCGCTCGGAATTCGCTGCGACAAGTACAAGGGGCGTAATCCCTGCCCACACGGGCGGAGC 962

Qy        939 G 939

Db        963 G 963


RESULT 15
BQ183029/c
LOCUS       BQ183029                656 bp    mRNA    linear   EST 15-JUL-2003
DEFINITION  UI-H-EU0-azs-e-03-0-UI.s1 NCI_CGAP_Car1 Homo sapiens cDNA clone
            IMAGE:5852906 3', mRNA sequence.
ACCESSION   BQ183029
VERSION     BQ183029.1  GI:20358579
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5852906"
                     /tissue_type="Osteoarthritic Cartilage"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Car1"
                     /note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_Car1 is a cDNA library containing the following
                     tissue(s): Osteoarthritic Cartilage The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_TISSUE=osteoarthritic cartilage
                     TAG_LIB=UI-H-EU0
                     TAG_SEQ=TGATCACGCT"
ORIGIN

  Query Match             29.5%;  Score 642.4;  DB 13;  Length 656;
  Best Local Similarity   99.8%;  Pred. No. 2.2e-113;
  Matches  643;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1534 GGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGG 1593

Db       651 GGGCCCACCTGAGGATCATCGACGGGGGCTCAGGCTACCTGTGTGAGATGGAGCCCGTGG 592

Qy      1594 CACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCA 1653

Db       591 CACACTTTGGCCTGGGGAAGGATGAAGCCAGCAGTGTGGAGGTGACGTGGCCAGATGGCA 532

Qy      1654 AGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACC 1713

Db       531 AGATGGTGAGCCGGAACGTGGCCAGCGGGGAGATGAACTCAGTGCTGGAGATCCTCTACC 472

Qy      1714 CCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCC 1773

Db       471 CCCGGGATGAGGACACACTTCAGGACCCAGCCCCACTGGAGTGTGGCCAAGGATTCTCCC 412

Qy      1774 AGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCC 1833
```

```
Db        411 AGCAGGAAAATGGCCATTGCATGGACACCAATGAATGCATCCAGTTCCCATTCGTGTGCC 352

Qy       1834 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 1893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        351 CTCGAGACAAGCCCGTATGTGTCAACACCTATGGAAGCTACAGGTGCCGGACCAACAAGA 292

Qy       1894 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 1953
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        291 AGTGCAGTCGGGGCTACGAGCCCAACGAGGATGGCACAGCCTGCGTGGGCTGGTGGAGCC 232

Qy       1954 CTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAAT 2013
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 CTGTGTTGAAGATAGTGACACCACAAGTTGGGAAGAGCCTTGGTCCCTGAATCACTGAAT 172

Qy       2014 CACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCA 2073
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 CACTGCCTTGAATCACCGCCTGGAATACCTGTTGATCAGGAACACTTACCTGGAACTTCA 112

Qy       2074 CTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAATACATTAAGATTTGGGGGTGC 2133
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 CTGAGCAGGATACAAACTTCTATTGTATTAAGCTATTAATACATTAAGATTTGGGGGTGC 52

Qy       2134 TACCTTACATAATAAATTCCCATTTCCTCTTGAAAAAAAAAAAA 2177
             |||||||||||||||||||||||||||||| ||||||||||||
Db         51 TACCTTACATAATAAATTCCCATTTCCTCTTGGAAAAAAAAAAA 8
```

Search completed: August 9, 2004, 21:15:51
Job time : 5112 secs